

# CHAPTER I

## Introduction

The game of tennis has become increasingly popular among participants and spectators around the world. As a lifetime activity, it offers challenges at every level, and both long and short term benefits for all ages. Many students across the United States are exposed to the game of tennis through their school's physical education program. However, tennis is only one of many sports offered. Detailed attention cannot be given to any one sport. Students have limited exposure, and the result is that few sports, if any, are mastered (Kelly, 1989).

Many physical education programs are equipped with county or state curricular guides. However, many of these curricular guides are outdated and should not be used by physical education teachers today (Anderson, 1988). In 1984, the United States Tennis Association (USTA), developed a Schools Program Curriculum specifically for physical education teachers to utilize in their programs. Since 1985, this curriculum has been adopted in 1,600 school districts across the country. Unlike most curriculums, the USTA Schools Program Curriculum is revised every two years in order to provide physical education teachers and students with the latest in tennis developments.

The USTA Schools Program Curriculum is designed to introduce students to the basic knowledge and fundamental skills of tennis. It also provides students with maximum practice opportunities for the basic strokes. It was developed so that students might receive all the lifetime benefits related to the sport. The program consists of eight progressions from striking with rackets and paddles to transitional games. Each progression includes important swing cues to help students remember specific aspects about the swing. A complete list of activities is included to provide multiple options for physical education teachers to use in assisting students to practice each progression.

One uncertain aspect about programs such as the USTA Schools Program Curriculum is how it determines what students have learned. Physical education programs traditionally use fitness testing as the only measure of formalized assessment (Buschner, 1994). Written objective tests and motor skill tests have also been recommended to assess what students have learned in physical education. However, research has shown that less than half of physical education teachers give written objective tests, while only one fourth never use tests of understanding (Hensley, Lambert, Baumgartner & Stillwell, 1987).

Many physical education teachers reject standardized skills tests and create their own for the following reasons (Buschner, 1994; Strand & Wilson, 1993; Veal, 1992): skills tests are considered to be invalid predictors of a student's playing ability; a student's performance of a skill is

inconsistent from day to day; and skills are typically examined out of context of the game (Veal, 1992).

Unfortunately, many physical education teachers do not assess their students' progress at all (Buschner, 1994). This usually occurs simply because physical education teachers often have twice the number of students as classroom teachers and time does not permit individualized assessment (Graham, 1992; Strand & Wilson, 1993). Others don't assess because they don't believe assessment will make a difference in students' lives (Buschner, 1994).

According to Hopple (1995), "the solution is not to discard assessment altogether, as unfortunately has happened in many situations, but rather to find worthwhile alternative ways to assess in the time you have. Performance tasks are among these alternative measures" (p.21).

Alternative assessments are currently being utilized in schools for several important reasons. First, there is a growing concern among teachers regarding the use of objective tests, e.g., multiple choice, matching, fill-in-the-blank, true/false, standardized sport skill tests and physical fitness tests (Hensley, 1997). It is now felt that these tests do not measure how well students can think, what students know about a subject, how sufficiently they can solve problems, and how responsible they are in helping to direct their own learning process (Brandt, 1993; Herman, 1992; Marzano & Costa, 1988; Mitchell, 1992; Pool & Bracy, 1992; Wiggins, 1987; 1989 a). Alternative assessments differ by challenging students on meaningful complex tasks. This is accomplished by requiring them to organize information they have learned, build on what they know, explain or arrive at some conclusion, and apply it in new situations. Alternative assessment methods can enhance learning, facilitate teaching, and according to Hensley (1997), result in greater student achievement.

Second, alternative assessments encompass methods for students to demonstrate their knowledge and understanding by creating, producing, or doing something, rather than just taking a written objective test (Herman et. al., 1992). Some of these methods include the following: exhibitions, interviews, observations, open-ended questions, and portfolios (Herman et. al., 1992; Hopple, 1997). Exhibitions enable students to choose their own ways to demonstrate what they have learned. Interviews give students opportunities to reflect on their achievement. Students can learn how to analyze and critique a movement by observing a skill. Open-ended questions challenge students to move beyond one right answer and explore other possibilities. Portfolios are collections of students' work over a period of time. "Higher - level assessment strategies such as portfolios, journals, and projects, challenge students to reflect on their learning, provide written analysis of skills and knowledge they have mastered, seek out new ways to use their skills and knowledge, and keep accurate records related to their learning" (Schiemer, 1996, p. 4).

Third, alternative assessments provide information concerning student learning in a specific program. Since alternative assessments are designed to be congruent with instruction and match a program, teachers then use the results of the assessments to determine if students really know and understand what they have been taught. Teachers can use the results of the assessment to help guide or improve their instructional efforts.

It is important for students in physical education to understand why and how they perform a motor skill. In addition to learning the history and rules of a particular sport, students need to be able to combine and organize what they have learned, explain and build on that information, and apply that information in various settings.

One specialized alternative method is authentic assessment. Authentic assessment evaluates student knowledge in a real life context through performance tasks that require student generated responses (Mustain, 1994). For an assessment to be considered authentic, the audience, content, constraints and purpose should be integrated in some form to real life situations and problems (Wiggins, 1989). Therefore, authentic assessments are designed to engage students on significant tasks that best represent real life experiences (Bruder, 1993). Authentic assessments in physical education enable teachers "to see the skill that is being evaluated. Connections of skills to real-life situations are used as learning indicators" (Veal, 1992, p. 90).

This thesis will show that outcome based learning goals can be accomplished through the use of alternative assessment. It will also show that it is possible to develop valid and reliable alternative assessment tasks to evaluate the knowledge of the physical education student.

### **Statement of the Purpose**

The purpose of this study was to develop two valid and reliable alternative assessment tasks based on the USTA School's Program Curriculum. These assessments might then be used to determine the effectiveness of the USTA School's Program Curriculum as taught in selected schools.

### **Research Questions**

The following research questions were used to guide this study.

1. Was it possible to develop satisfactory levels of validity on each alternative assessment task based on the USTA Schools Program Curriculum?
2. Was it possible to develop satisfactory levels of reliability on each alternative assessment task based on the USTA Schools Program Curriculum?

### **Operational Definitions**

Experienced University Varsity Tennis Players - Those who play on the varsity tennis team at the university level.

Experienced University Tennis Players - Those who participate in a regular basic instruction physical education class and based on a questionnaire,

referring to tennis knowledge and experience, indicate experience and a former member of their high school tennis team.

Inexperienced University Tennis Players - Those who participate in a regular basic instruction physical education class and based on a questionnaire referring to tennis knowledge and experience, indicate no experience or limited experience with the sport.

Reliability - making sure that the score is justifiable and accurate, and whether test scores remain consistent over time despite changes in the assessment process (Brand, 1993).

Inter-Rater Reliability - The degree to which different raters can obtain the same scores on the same tasks (Thomas & Nelson, 1990).

Intra-Rater Reliability - The degree to which one rater remains consistent over time (Herman et. al., 1992).

Subject Reliability - The degree to which subjects remain consistent over time (Herman et. al., 1992).

Validity - indicates whether the alternative assessment scores accurately reflect the knowledge, skills, and abilities they are intended to measure (Herman et. al., 1992).

Construct Validity - students with known differences in ability or understanding complete assessment; if scores on the assessment are different then construct validity is inferred (Thomas & Nelson, 1990).

Content Validity - assessment tasks measure the content covered in the USTA Schools Program Curriculum (Thomas & Nelson, 1990).

### **Assumptions**

The following assumptions were made for this study:

1. Subjects in this study were honest in completing a questionnaire stating their experience with the game of tennis.
2. Subjects chosen for this study tried their best on each of the alternative assessment tasks.

### **Limitations of Study**

The subjects in this study were experienced tennis players and inexperienced tennis players at the university level.

### **Delimitations of Study**

The delimitation's for this study were as follows:

1. Only university students were used as subjects for this study.
2. Only experienced tennis players and inexperienced tennis players were selected as subjects for this study.
3. Only two alternative assessment tasks were developed and used for this study.
4. Only 18 subjects participated in this study.
5. Testing environments were not consistent between groups.

### **Research Hypothesis**

The following results were hypothesized for this study:

1. Experienced university varsity tennis players will score better on both alternative assessment tasks as compared to experienced university tennis players and inexperienced university tennis players.
2. Experienced university tennis players will score better on both alternative assessment tasks as compared to inexperienced university tennis players.
3. It is possible to develop valid and reliable alternative assessment tasks based on the United States Tennis Association Schools Program Curriculum.

### **Significance of Study**

One major area of significance for this study lies in the ability of alternative assessment to provide accountability for the methods of instruction and the learning that takes place as a result (Hensley, 1997). Many physical education programs lack accountability for what students have learned. For example, students enrolled in a physical education class are traditionally graded on class attendance, dressing out for physical education, and participation (Imwold, Rider, & Johnson, 1982). As a result, the completion of the class is the only thing for which students in physical education are typically held accountable (Lund, 1992).

This differs significantly from the types of accountability required of teachers in academic subjects. For example, math and English teachers are held accountable for students' performance on minimal competency exams. If students pass these exams, they will then be promoted to the next level in English or math.

Historically, physical education teachers have used fitness testing as the only measure of formalized assessment in physical education. Some physical education programs use standardized skill testing as a major part of their program. Less than 50% of physical education teachers use objective written tests (Buschner, 1994). Unfortunately, due to the excessively high numbers in physical education classes today even these forms of assessment and accountability have fallen by the way side.

Alternative assessment methods enable teachers to document student learning throughout a unit as well as a program, and can prove to be a way of lending credibility to physical education programs. It is also a way to hold students and teachers accountable for subject matter competency rather than simply the completion of the class (Lund, 1992). Since many physical education programs lack accountability, being able to document student learning is crucial. Physical education has been at the bottom of the education hierarchy in the past "partly because physical educators have not been successful in convincing tax payers, parents, and administrators of its importance" (Wood, 1996, p. 21). If teachers were to document the profiles of students knowledge, skill and fitness levels, and improved skill levels, they would have some data to support the value of their programs. This is especially important when it comes time for program review, budget cuts, and school restructuring (Lund, 1992). By

using alternative means of assessment such as performance tasks, teachers can demonstrate to administrators, parents, the community and themselves that their students are truly learning the course material and meeting program outcomes (Hopple, 1995). Assessment results can also be used to justify the need for more faculty, new equipment, or curriculum changes. (Freeland & Moore, 1996)

Alternative assessment is being utilized programatically in state-wide programs. The state of Kentucky, for example, is using alternative assessments to measure school improvement (Lund, 1992). Students can no longer just complete the class and receive a grade (Lund, 1992). They must be able to demonstrate proficiency as well as mastery of the subject before the school is released from the responsibility of educating the student (Lund, 1992).

Alternative assessment in physical education requires teachers to use assessment to aid students' learning, not just to test students and assign them a grade. This method tests students on their ability to demonstrate their knowledge in a way that is meaningful to them.

Another area of significance for this study is that alternative assessments can be designed to match a program, thereby providing ownership. For example, physical education teachers can test their students using alternative assessments they have designed specifically to measure program outcomes. These are of more value than the skill tests frequently used which measure skills out of context of the game (Veal, 1992). Alternative assessments are also more meaningful than written objective tests which primarily measure factual knowledge. This is true because written tests cover information that is memorized the night prior and often forgotten soon after (Wiggins, 1992). Even though these tests draw out important distinctions from students, the tests do not measure a student's ability to combine and organize information in various settings. They are usually tests that have been used repetitively and are not updated to match the content of the program.

A final area of significance for this study is that few studies rarely address alternative assessment procedures pertaining to physical education. Further research of alternative assessment and how it can be adapted for physical education programs would help lend credibility to assessment in physical education. The use of these assessments will hopefully increase the accountability and credibility for physical education teachers and their programs.

## **Summary**

Alternative assessments assess student knowledge through performance tasks that require student generated responses (Mustain, 1992). They are being recognized as alternative methods to standardized skill tests and written objective tests in assessing what students have learned. In physical education, alternative assessments could increase the accountability for teachers and lend credibility to the field. They can be used for program assessment, provide ownership to a program and alternative assessments might also provide the hard data on learning needed to convince those of budget and finances to lend more support and respect to physical education programs. This study attempted to produce two valid and reliable alternative assessment tasks based on the USTA Schools Program Curriculum and determine if alternative assessment would be valuable for physical education.

## **CHAPTER II**

### **Review of Literature**

The purpose of this study was to develop two valid and reliable alternative assessment tasks based on the USTA Schools Program Curriculum. The following areas of related literature were identified as being applicable to review: 1) tennis in schools, 2) assessment in physical education, 3) limitations of traditional educational assessment, 4) new face of assessment, 5) characteristics of alternative assessment, and 6) what alternative assessment look like in physical education.

#### Tennis in Schools

Since the early 1970's tennis increased in popular among participants of all ages. In 1968, an estimated population of eight million people were playing tennis. A decade later, estimates increased as high as 40 million (Kraft, 1979). The National Federation of State High School Associations (1974) estimated that boys' and girls' high school tennis teams have also boomed. In 1968, 83,717 boys were participating on 6,221 tennis teams, and in 1978, twice as many boys were actively playing tennis on 8,889 teams. Girls teams experienced even more growth with an explosion to 147,365 participants in 1978 from 2,649 in 1970.

In addition to tennis teams, many schools offer tennis as part of their physical education curriculum. In 1984, the United States Tennis Association (USTA) developed a schools program curriculum specifically for physical educators to utilize in their programs (USTA, 1993). In 1993, over 1,600 school districts across the country had adopted this curriculum into their schools (USTA, 1993). The USTA realizes that physical education classes encompass large numbers of students and that tennis is only one of many sports offered. Therefore, their primary goal in designing the program is for physical education teachers to introduce students to the basic knowledge and fundamentals of tennis so that they can enjoy the benefits that accumulate from playing tennis for a lifetime (USTA, 1993).

The USTA curriculum contains eight skill progressions including introduction to striking with rackets, basic forehand skills, basic backhand skills, forehand and backhand combination skills, the simplified serve, introduction to the volley, serve and return of serve, and transitional games (USTA, 1993). Each progression offers teachers a variety of activities that can be used for all skill levels and swing cues that would help students remember important aspects about the swing. The USTA currently mirrors the assessment methods in other physical education programs by recommending motor skill tests and written objective tests to assess student knowledge and skill levels.

#### Assessment in Physical Education

Some physical education programs have assigned grades to students for attendance, attitude, dressing out, participation, and for taking showers (Imwold, Rider & Johnson, 1982; Strand and Wilson, 1993). While these factors may be important to the administration of a class, they do not provide information about what students have learned or achieved.

Physical education programs have traditionally used fitness testing, motor skill tests, and written objective tests to assess students motor performance and knowledge in physical education (Buschner, 1994; Strand & Wilson, 1993). In some programs, fitness testing provides the only measure of formalized assessment (Buschner, 1994). One major criticism of fitness testing is that physical educators do not see fitness as part of the whole educational process for students (Symons & Gascoigne, 1990). Students must be able to see their scores, understand and interpret them, establish goals based on their scores, and evaluate their progress toward reaching these goals (Hopple, 1994).

Standardized skill tests are typically used to assess the motor performance of students (Buschner, 1994; Strand & Wilson, 1993). Skill tests can assist teachers by providing information on whether the class objectives were achieved and identify the strengths and weaknesses of the teacher's instruction. However, many physical education teachers do not use preconstructed skill tests, and criticize them for the following reasons. First, they do not match the goals of their program (Strand & Wilson, 1993; Veal, 1992). Second, the skill tests are considered invalid predictors of students' playing ability (Avery et. al., 1979). Third, student's performance of a skill is inconsistent from day to day (Veal, 1992). Finally, skills are typically examined out of context of a game (Strand & Wilson, 1993; Veal, 1992).

In tennis, the skills tests most commonly used to assess students physical performance are those which assess the ground strokes and the serve (Avery et. al., 1979). Unfortunately, most of these tests do not examine skills in game-like situations. The majority of these skill tests were developed and published in 1938 through 1981 and have never been updated. Tennis skill tests that have been used to determine the serving ability of students include those developed by Cobane (1962), DiGennara (1969), Driver (1973), Hewitt (1966), and the National Tennis Foundation, Inc. (1974). Four of these skill tests allow students to serve only one time per trial, whereas in game-like situations, students would have two trials to put the ball in play. These tests also require students to serve from only one of the two service courts, as compared to the game where a student would serve from both service courts. Because of these criticisms, teachers either fail to use skill tests or they create their own (Strand & Wilson, 1993; Veal, 1992). According to Veal (1992), teacher-made tests are perceived as better tests because they fit the context" (p. 90).

Written objective tests may also be used in physical education programs to assess students' knowledge of a sport (Buschner, 1994). However, research has shown that less than half of physical education give written

objective tests, while 25% never use tests of understanding at all (Hensley, Lambert, Baumgartner, and Stillwell, 1987). Written objective tests are traditionally given at the end of a unit and have usually measured a student's ability to memorize factual information such as the history, safety, rules, equipment, and etiquette of a sport. This type of test is important in measuring the analytical intelligence of students, but it is also relies on the kind of information that is easily forgotten (Wiggins, 1992). According to Strand and Wilson (1993), and Veal (1992), for students to achieve mastery levels of learning, assessment should be ongoing, and throughout a unit (formative), not just at the end of the unit (summative).

Many physical education teachers do not assess their students' progress at all, for two main reasons. First, physical education teachers typically handle twice the number of students taught by their counterparts in the classroom setting. The high volume of students in their classes imposes time constraints that make it nearly impossible to perform assessments (Strand & Wilson, 1993; Graham, 1992). Second, many physical education teachers do not think assessment will make a difference in students lives (Buschner, 1994).

#### Limitations of Traditional Educational Assessments

Educational assessment in schools today is in the process of undergoing changes and reform. Alternative assessment methods are exploding in popularity, while old methods, such as traditional standardized testing and multiple choice tests, are being seriously questioned (Brandt, 1993; Herman, 1992; Madaus & Kellaghan, 1993; Shepard, 1989; and Sternberg, 1989). Many educators are wondering if the test scores on standardized tests represent significant learning outcomes and if improvements in test score performance actually represent improvements in learning (Brandt, 1993; Herman, 1992; Madaus & Kellaghan, 1993; Shepard, 1989; Sternberg, 1989). In the 1970's, standardized tests were thought to be reasonable indicators of student learning (Shepard, 1989). However, today the criticisms of standardized tests are abundant and center on their limitations (Brandt, 1993; Herman, 1992; Madaus & Kellaghan, 1993; Shepard, 1989; Sternberg, 1989).

One criticism points out that the content of these tests is negotiated among curriculum experts and teachers. The publishers do an analysis of content to make sure the content matches the objectives from widely-used textbooks, thus limiting the scope and the complexity of the content covered (Shepard, 1989). Another concern focuses on the construction of the tests limited by the emphasis of basic skills. This emphasis limits content and narrows the curriculum (Madaus & Kellaghan, 1993). The 1992 update issue of Association for Supervision and Curriculum Development stated that "Exams commonly used to measure how well students are doing - standardized, norm-referenced, multiple choice tests of basic skills of knowledge - drive classroom practice in the wrong direction, reinforcing a fragmented curriculum that fails to teach students how to solve complex problems or think for themselves" (p.3).

The third limitation of standardized tests is that they measure lower-order thinking, the ability to recite memorized factual information (Newmann & Wehlage, 1993). Students receive bits of information to be memorized rather than understood. They spend most of their time preparing for and taking tests that focus on their analytical intelligence, which has little significance in real-world situations (Sternberg, 1989). Marzano and Costa (1988) have found that "Performing well on standardized tests has little to do with students' thinking abilities and a great deal to do with how well they have learned factual knowledge represented in test items" (p.67). If students are to learn anything, they must be able to demonstrate their abilities to think and to apply their knowledge productively. This is not accomplished by just filling in the correct answer on standardized tests (Mitchell, 1992). Traditional forms of assessment such as paper and pencil or motor skill tests "tend to measure narrowly defined cognitive, affective, and psychomotor objectives, treat behavioral domains separately, and often fail to measure performance under realistic or authentic conditions" (Wood, 1996, p. 20).

A fourth criticism of traditional standardized tests is that teachers have been pressed to "teach to the test" because of the high stakes associated with standardized tests (Madaus & Kellaghan, 1993; Shepard, 1989). They will avoid other relevant material in order to prepare students to take the test. Test scores may go up, but there is no gain in achievement. As a result, important learning goals are omitted, the test results are not valid measures of the true knowledge a student possesses, and the tests diminish the meaning of the tests results (Shepard, 1989).

Finally, written objective tests, such as multiple choice tests, also lay down the format of the tests in ways that limit the content (Shepard, 1989). These tests are primarily used because of the high volume of students taking the tests and the time constraints involved (Shepard, 1989). Even though multiple choice items can indicate important distinctions between students, the tests do not measure a student's ability to combine and organize information, explain and build on that information, and apply that information in various settings. Furthermore, the majority of educational outcomes that educators value for students cannot be translated into objective paper and pencil tests (Stiggins, 1991).

#### New Face of Assessments

Alternative assessments have grown out of concern among teachers that standardized tests and written objective tests do not measure how well students can think, what students know about a subject, how sufficiently they can solve problems, and how responsible they are in helping to direct their own learning process (Brandt, 1993; Cronin, 1993; Herman, 1992; Marzano & Costa, 1988; Mitchell, 1992; Pool & Bracey, 1992; Wiggins, 1989a, b). Another concern is that standardized tests do not precisely measure the capabilities of students from disadvantaged homes, students

with learning disabilities, students who cannot speak English well, and those who have difficulty coping in the mainstream (Pool & Bracey, 1992).

Alternative assessments are characterized as performance-based assessments. They take on the form of tasks instead of tests. The Association for Supervision and Curriculum Development states that "In the world we have in mind, you don't give kids tests, you give them tasks. And in giving them tasks, you aren't taking time away from classwork, that is classwork" (O'Neil, 1992, p.3). The result of student work is termed a product, and scoring rubrics or guides are developed and designed to match the content of the task and used to evaluate students products.

One specialized type of alternative assessment is authentic assessment. Authentic assessment assesses student knowledge in a real-life context through performance tasks that require student generated responses (Mustain, 1994). Authentic assessment requires students to apply their knowledge to significant tasks that best represent the experiences they will face in real-life. Veal (1992), refers to authentic assessment as the "direct examination of students on intellectual tasks rather than on indirect proxy items such as those found on standardized tests" (p.89). Madaus and Kelleghan (1993) define authentic assessment as the "direct assessment of complex performances" (p. 459).

#### Characteristics of Alternative Assessment

Wiggins (1989a, b) states that a true test requires the performance of exemplary tasks. Authentic forms of alternative assessments attempt to replicate the challenges and standards of performance that students will typically face in the real world (Wiggins, 1989a, b). Students are engaged in meaningful tasks that are significant to their achievement as compared to memorizing facts that have no value in later life.

Authentic assessments require a real audience and do not rely on unrealistic or arbitrary time constraints. The assessment task is known ahead of time to students as compared to traditional tests. Authentic assessments require some collaboration with others. Real-life problems often require people to work together. Herman (1992) states that "groups may facilitate learning by modeling effective thinking strategies, scaffolding complicated performances, providing feedback and valuing elements of critical thought" (p.75). Alternative assessments including those that are authentic make assessment and feedback central to students. The assessments are ongoing, and occurring over and over again, giving students ample opportunities to practice, rehearse, and retake (Veal, 1992; Wiggins, 1989a, b). According to Wiggins (1989a, b) students become better educated by taking the tests over and over.

Alternative assessments attempt to engage students in higher-order thinking (Mitchell, 1993; Newmann & Wehlage, 1993; Schnitzer, 1993; Wiggins, 1989a, b, 1992). Higher-order thinking requires students to manipulate information in ways that provide new meanings and implications. Students combine facts and ideas in order to synthesize,

generalize, explain, hypothesize, or arrive at some conclusion. By using these processes to manipulate information, students can solve problems, and develop new meanings and understandings (Newmann & Wehlage, 1993).

Alternative assessments are characterized as complex tasks that challenge students by asking them to create, produce, or do something (Herman et. al., 1992). Authentic forms of alternative assessment require the student to appropriately apply what has been learned to real-world situations and contexts. When students can relate their knowledge to something that is real, learning becomes more meaningful. Alternative assessments incorporate a depth of knowledge more than breadth (Wiggins, 1989a). Depth is defined as covering few topics in systematic and connected ways. Students make clear distinctions, develop arguments, solve problems, and construct explanations (Newmann & Wehlage, 1993). Standardized testing and written objective tests primarily use content that is thin or superficial and does not deal with the significant concepts of a topic or discipline (Wiggins, 1989a).

Alternative assessments involve criteria that assess students on what is most important, not what is easiest to score. Students know the criteria and the goal of the task prior to tackling it. Alternative assessments also use a multifaceted scoring system instead of assigning a grade.

Alternative assessments are tailored to student learning styles, aptitudes, and interests. Tasks can be attempted by all students. Unlike multiple choice and standardized tests, alternative assessments do not have only one correct answer. Students are mentally challenged to explore open-ended questions and problems by drawing on their own inferences.

#### Alternative Assessment in Physical Education

Alternative assessments are being recognized as another way to assess student learning and improve teacher instruction in physical education (Hensley, 1997; Hopple, 1995, 1997; Joyner & McMannis, 1997; Lund, 1992, 1997; Melograno, 1997; Veal, 1992). Alternative assessments are also a means of holding teachers and students accountable for knowing and understanding physical education rather than simply providing a grade for the completion of the class (Lund, 1992). Lund (1992) states that alternative "assessments are a way to hold students accountable for knowing how the pieces fit together and for demonstrating command of a subject matter or competence in a skill. It is not enough to pass skill tests in physical education. Students must be able to integrate skills into games, dances and so on to demonstrate mastery" (p. 358).

Veal (1992) describes alternative assessment in physical education as "regular and ongoing" (p. 90). Instead of just assessing students at the end of a unit, alternative assessment could take place throughout a unit. For example, in a tennis class, students could document their progress on

individual skills such as the tennis serve from the beginning to the end of class.

Another feature of alternative assessment in physical education described by Veal (1992) is the "connection between daily instructional tasks and assessment" (p. 90). According to Hopple (1995, 1997), and Wiggins (1989a), teachers should first decide what is important for students to know. When teachers have defined what they want students to know and do, they can then design an alternative assessment not only to match the goals of their program, but also to match their instruction as well.

An important feature of the authentic form of alternative assessment in physical education is that teachers can "see the skill" that is being assessed in a real-life context (Veal, 1992). For example, teachers can assess students performance in a game of tennis and collect information such as percentages of serves. Students can also be taught to collect this information which would link learning in physical education with other disciplines such as mathematics (Veal, 1992).

Being able to explain and demonstrate a skill, sequence, dance etc., is another feature of alternative assessment in physical education. For example, students enrolled in a physical education class may be required to explain the critical components of a skill, why the critical components are important, and demonstrate those components on videotape. An alternative assessment in fitness might require students to explain and calculate a target heart rate and, based upon their results, write and explain how to get in shape or explain ways to stay in shape (Lund, 1992).

The literature has indicated that alternative assessments is another way to assess student learning in physical education (Hensley, 1997; Hopple, 1995, 1997; Joyner & McMannis, 1997; Lund, 1992, Melograno, 1997; Veal, 1992). Standardized skill tests and written objective tests could be used as adjuncts to alternative assessments. Skill tests could serve as formative evaluations to provide baseline information to teachers on students learning to master a skill, instead of using them as summative evaluations. Giving skill tests throughout units would enable teachers to give constant feedback to students, thus enabling them to improve their skill levels.

### **Summary**

This review of literature focused on six main areas of literature relevant to the purpose of this study. The first area described tennis in schools. Research in this area states that each year tennis continues to gain popularity as a lifetime sport among participants (Kraft, 1979). In 1984, the USTA developed a schools program curriculum containing eight skill progressions from striking with rackets to transitional games specifically for physical educators to utilize in their programs.

The second area described assessment in physical education. Traditionally, some physical education programs have used fitness testing, standardized skill tests, and written objective tests to assess students' motor performance and knowledge in physical education (Buschner, 1994; Strand & Wilson, 1993). Many physical education teachers do not assess their students' progress at all because of the high volume of students, the lack of time, and the disbelief that assessment will make a difference in students' lives (Buschner, 1994; Graham, 1992; Strand & Wilson, 1993).

The third area described the limitations of educational assessment. The first limitation is that the content of standardized tests matches the objectives from widely used textbooks, thus limiting the breadth and depth of the content covered (Shepard, 1989). Second, the construction of the test is limited by the emphasis on basic skills, thus limiting the scope of the content and narrowing the curriculum (Maddaus & Kellaghan, 1993). Third, standardized tests primarily measure lower-order thinking (Newmann & Wehlage, 1993). Fourth, because of the high stakes associated with standardized tests, teachers have been pressed to "teach to the test," therefore omitting important learning goals and diminishing the meaning of the test results (Shepard, 1989). Finally, written objective tests, such as multiple choice tests, also lay down the format of the tests in ways that limit the content (Shepard, 1989).

The fourth area described the new face of assessments today. Alternative assessments are characterized as performance-based assessments. One specialized type of alternative assessment is authentic assessment. Authentic assessment assesses student knowledge in a real-life context through performance tasks that require student-generated responses.

The fifth area described the characteristics of alternative assessment. Some of the characteristics included: (1) engaging students in meaningful tasks that are significant to their achievement (Wiggins, 1989a, b, 1992), (2) requiring a real audience and not relying on unrealistic or arbitrary time constraints in the authentic form of assessment (Wiggins, 1989a, 1992), (3) making assessment and feedback central to students in higher-order thinking (Mitchell, 1993; Newmann & Wehlage, 1993; Schnitzer, 1993; Wiggins, 1989a, 1992), and (4) involving criterion that assesses students on what is important, not what is easiest to score.

The final area described various ways that alternative assessment could be used in physical education. Veal (1992) describes the authentic form of assessment in physical education as regular and ongoing, a connection between daily and instruction tasks, and permitting teachers to "see the skill" that is being assessed in a real-life context (p.90). This area also mentioned some ideas on how alternative assessments could become another way of holding teachers and students accountable in physical education.

## **CHAPTER III**

### **Methodology**

This study was designed to develop two valid and reliable alternative assessment tasks that might be used in conjunction with the United States Tennis Association Schools Program Curriculum. These assessments could be used to determine what students have learned as a result of participation in the program.

#### **Setting**

This study was conducted at a state university located in a large southeastern state. The state university is located in a rural region of the state and consists of approximately 23,000 students. The university's tennis team is a member of the National Collegiate Athletic Association (NCAA). The athletic department offers eight full scholarships for the female tennis team and six full scholarships for the male tennis team.

The university offers a Basic Instruction Program to all students. The courses offered are equivalent to one credit hour and are graded on the basis of pass/fail only. Some of the courses included in the program are aerobic dance, badminton, bowling, cycling, jogging, modern dance, racquetball, tennis, swimming, volleyball and weightlifting.

#### **Subjects**

The subjects for this study were experienced university varsity tennis players (EUVTP's), experienced university tennis players (EUTP's) and inexperienced university tennis players (IUTP's). EUVTP's were defined as those who played on the school's varsity tennis team. EUTP's were defined as those who had tennis experience and were a former member of their high school tennis team. IUTP's were defined as those who had limited or no playing experience of the game of tennis.

Prior to the selection of subjects for this study, a consent form was developed and distributed to prospective subjects. Subjects who returned the consent form became a part of the "pool" from which subjects were chosen (Appendix A).

Six EUVTP's were selected for this study based on the following criteria: (1) Varsity team affiliation and (2) the willingness to complete two or more assessment tasks.

Six EUTP's were selected for this study based on the following criteria: (1) participation in a regular basic instruction physical education class, (2) the willingness to complete two or more assessment tasks, and (3) ranked highest on questionnaire results of tennis knowledge and experience. Each response on the questionnaire had been given a numerical value,

the highest being 28 and the lowest being 7. Students who scored 20 or higher became part of the pool from which subjects were chosen for this study (Appendix B).

Six IUTP's were selected for this study based on the following criteria: (1) participation in a regular basic instruction physical education class, (2) the willingness to complete two or more assessment tasks, and (3) ranked lowest on questionnaire results of tennis knowledge and experience. Students completed the same questionnaire as EUTP's. Students who scored lower than ten became part of the pool from which subjects were chosen for this study.

### **Instruments**

Two alternative assessment tasks were developed for this study based on the USTA School's Program Curriculum. The following section describes the development of Task One and Task Two as well as the scoring rubrics that were used to judge subject performance on both tasks.

#### **Task One: Creating a video of basic tennis skills**

According to Herman, Aschbacher, and Winters (1992), the key to a good assessment is matching the assessment task to the intended student outcomes. The major outcome of the USTA's School Program Curriculum is for students to enjoy the short and long term benefits of playing tennis. Some of the short and long term benefits are as follows: tennis as a lifetime sport, thinking skills such as ball placement strategies, recreational benefits (playing for the sheer pleasure of the game), mathematic skills (learning how to keep score and knowing geometrical terms such as angles and lines), communication skills (being able to explain a skill or concept as well as perform it), social skills (constant interaction with partners and opponents), and finally, emotional development (players must be self reliant and be their own officials) (USTA, 1993).

The purpose of Task One was to see if students could apply the knowledge described in the USTA Schools Program by developing a video of basic tennis skills. The results from subject products could then be used to determine the effectiveness of the USTA Schools Program Curriculum. Creating the task description for Task One involved three steps: defining the outcome(s) intended for the assessment, describing the assessment administration process, and developing the actual question, prompt or problem (Herman et. al., 1992). Since one of the outcomes of the USTA Schools Program Curriculum was for students to explain a skill or concept as well as perform it, then the first step in the task description for Task One was defining the outcome intended for measurement as the subjects' ability to explain and demonstrate basic tennis skills. See Table One.

The second step in creating a task description was the assessment administration process (Herman et. al., 1992). This step described how Task One was administered to students, whether students worked independently or in groups, the materials or equipment needed to

administer the task, and the time allowed to complete the task. For Task One, subjects were administered the task at the university tennis courts or in the university gym. Subjects worked together in groups of three to complete the task. Poster board, markers, tennis rackets and tennis balls were available to subjects. A time period of 45 minutes was allotted to subjects to complete this task.

The final step in creating a task description included the format of the task, identification of the audience for which the task was designed, and a description of the directions students would use to complete the task (Herman et. al., 1992). For Task One, subjects created a video of basic tennis skills for children. Working in groups of three, subjects chose three different basic tennis skills. Each group then wrote at least four cues pertaining to the skill they chose on a posterboard. When subjects were finished writing on the posterboard, they were ready to produce their video. The researcher videotaped subjects as they explained the skill they chose. Subjects could refer to their poster during their explanation of the skill while they or someone in their group demonstrated it.

**Table 1**

**Task One Description**

---

<b>Outcome to Be Measured</b>	Subject's ability to explain basic tennis skills.
<b>Assessment Administration Process</b>	
a. Administration of task	University tennis courts or university gym
b. Group or individual roles	Groups of three
c. Materials / equipment	Posterboard, markers, tennis racket/s, tennis balls, video camera, video tape
d. Time allowed	45 minutes
<b>Actual Question, Prompt or Problem</b>	
a. Format of task	Creation of a video of basic tennis skills
b. Audience	Children
c. Directions for task	<ol style="list-style-type: none"><li>1. Form working groups of 3 subjects per group.</li><li>2. Choose 3 different basic tennis skills per group.</li><li>3. Write at least 4 cues pertaining to each skill selected on a posterboard.</li><li>4. Explain and demonstrate each skill while being videotaped.</li></ol>

## Task Two: Developing a booklet of basic tennis skills

The purpose in developing Task Two was to see if students could apply the knowledge described in the USTA Schools Program by developing a booklet of basic tennis skills for high school students to utilize to teach children the basic skills of tennis. Creating the task description for Task Two involved three steps: defining the outcome(s) intended for the assessment, describing the assessment administration process, and developing the actual question, prompt, or problem (Herman et. al., 1992).

The first step defined the outcome intended for measurement in Task Two as the demonstration of subjects' knowledge of the basic skills of tennis. See Table Two. The second step in creating a task description was the assessment administration process (Herman et. al., 1992). This step described how Task Two was administered to subjects, whether subjects worked independently or in groups, the materials or equipment needed to administer the task, and the time allowed to complete the task. Subjects were administered Task Two at the university tennis courts or in the university gym depending on weather conditions. Subjects worked independently to develop a booklet. To administer this task, subjects were given writing utensils, markers, a folder to design a cover for the booklet and the skill pages that made up the booklet. Subjects were allotted 45 minutes to complete this task.

The final step in creating the task description for Task Two included the format of the task, identification of the audience for which the task was designed, and a description of the directions needed to complete the task (Herman et. al., 1992). For Task Two, subjects developed a booklet of basic tennis skills to which high school students could refer to when teaching children the basic strokes of tennis at a summer recreation center. Directions for the development of the booklet included the following: subjects were to identify the four basic skills to begin tennis play, list three cues of tips for each skill, and describe three ways to practice each skill. Subjects were also asked to design a cover for their booklet.

**Table 2**

**Task Two Description**

---

<b>Outcome to Be Measured</b>	Subjects demonstration of their knowledge of the basic skills of tennis.
<b>Assessment Administration Process</b>	
a. Administration of task	University tennis courts or university gym
b. Group or independent roles	Independent
c. Materials / equipment	Writing utensils, markers, paper to be used to design a cover for the booklet, skill pages that made up the booklet
d. Time allowed	45 minutes
<b>Actual Question, Prompt or Problem</b>	
a. Format of task	Develop an instructional booklet of basic tennis skills.
b. Audience	High school students teaching children the basic skills of tennis strokes at a summer camp.
c. Directions for task	1. Identify the four basic strokes needed to begin tennis play.  2. List three cues or tips for each skill.  3. Describe three ways to practice each skill.  4. Design a cover for the instructional booklet.

Scoring Criteria/Rubrics

Scoring criteria/rubrics were developed by the researcher and utilized for judging student performance on Task One and Task Two (Appendix E & F). Standards were also set for what was an accepted performance (Herman et. al., 1992). Scoring criteria for Task One and Task Two was based on the goals and objectives of the USTA School's Program Curriculum (1993). According to Herman, Aschbacher, & Winters (1992), good criteria often include both descriptive and numerical values. The rubrics for this study contained descriptions that served as the basis for judging student performance (Herman et. al., 1992).

Additionally, the rubrics contained a scale of values on which to rate each dimension (Herman et. al., 1992). The scales of values for this study used a four-point rating. Performance was rated numerically, but each numerical score was attached to an evaluation ranging from "beginner" to "advanced". For this study, a score of "4" represented those subjects who characterized an advanced level of tennis knowledge; a score of "3" represented those characterized an intermediate level of tennis knowledge, a score of "2" represented those who characterized an advanced beginner level of tennis knowledge, and a score of "1" represented those who characterized a beginners level of tennis knowledge.

### **Determining Reliability**

The purpose of determining reliability was to ensure that subjects' scores on Task One and Task Two were not the result of doubtful judgment and to develop consistency between raters and subjects (Herman et. al., 1992). A rater training program was developed in order to create consistent, reliable scoring procedures. This section describes the development and implementation of the rater training program for Task One and for Task Two. It also explains three methods of determining reliability that were used for this study. Two relate to raters and one to subjects.

#### **Rater Training - Task One**

For Task One, one rater and the researcher were used to ensure that subjects' scores on Task One were not the result of doubtful judgment and to develop consistency among the rater and the researcher. The rater was a doctoral candidate in physical education and had an extensive background in teaching physical education and tennis particularly at the collegiate level. It was evident from the background experience of the rater, that she had the content knowledge concerning the basic skills of tennis. Also, prior to this study, the rater and the researcher had taken a university graduate course in Authentic Assessment. Hence, the rater and the researcher had an understanding of authentic assessment, authentic assessment tasks, scoring rubrics, and reliability and validity issues.

To begin the rater training program, the rater received a copy of Task One and was asked to silently read through the assessment task. After reading the task, the rater had an opportunity to ask questions concerning Task

One. The following information was also discussed: the topic of the task, what the results could be used for, the subjects completing the task, the place where subjects would complete the task, the time allotted to subjects to complete the task, and the administration of the task.

General directions for Task One were also reviewed so that the rater would have a clear understanding of what was expected of subjects. For example, (1) subjects worked together in groups of three, (2) each subject in a group selected a different basic tennis skill, (3) each subject explained the skill they chose and demonstrated the skill or asked someone else in their group to demonstrate, (4) each subject wrote at least four cues or tips pertaining to the skill they chose on a piece of posterboard, and (5), each subject had four minutes to explain the basic tennis skill on video. Each skill chosen should include a complete demonstration to go along with the explanation of the skill.

The rater was then given a copy of the scoring rubric for Task One and was provided with a clear explanation of the scoring criteria by the researcher. Criteria dimensions and scale values were defined and examples of scoring criteria were available for her to view. She was informed that the following skills of tennis were considered acceptable answers for Task One according to the USTA Schools Program Curriculum (1993): forehand groundstroke, backhand groundstroke, serve, forehand volley, and backhand volley. The drop shot, lob, smash, slice etc., would not be accepted because the USTA recognizes these skills as "advanced or specialty shots/skills" not "basic" skills. There were many acceptable critical components for each skill. Most of the time the critical components subjects used had similar meanings to those listed on the chart but might be identified by a different name. A chart listing acceptable critical components (cues) developed by the researcher and based on the USTA Schools Program Curriculum was available for the rater to use (Appendix G). However, the rater was reminded that there could be more additional acceptable answers not found on the chart. If the rater and the researcher agreed to add some appropriate critical components to some of the skills, they could do so during rater training. Figures from the USTA Schools Program Curriculum (1993, pgs. 79, 80, 82, 83) illustrating proper stroke analysis were also available to help ensure consistency. The rater could refer to the stroke analysis figures if she had any questions concerning a subject's skill demonstration while scoring examples or real products.

Prior to scoring sample products, the rater and the researcher viewed 8 sample training products that were previously scored by the researcher. Each value on the scoring criteria was represented twice in all sample training products. The training products were also representative of the entire range of subject performance likely to be encountered in the actual scoring process. A clear justification for each rating was explained and discussed by the rater and the researcher. Practice scoring during rater training provided the rater with immediate feedback about her judgments and ample opportunities for her to ask questions. As the rater became

more proficient with the scoring process, she had opportunities to score additional sample training products and discuss borderline products. The rater also came to understand that her job was to make a judgment based on the scoring criteria and not her own inclinations.

The rater then scored 11 sample products that were previously scored by the researcher from several pilot studies conducted prior to this study. The rater scored one product at a time. She was provided with paper to take notes (write down critical components) concerning subject products. The video segment could be played several times during rater training (as well as during the actual scoring process) until the rater was ready to move on to the next segment. For example, the rater could write down the critical components she heard first, then replay the segment again to view the demonstration and check to see if the cues were demonstrated correctly.

Each product from the pilot study had been given a number. Rater 1 had her own score sheet for reporting scores which was used during rater training (Appendix H). The rater viewed the product, took notes, individually scored the product, and then wrote down the score in the box adjacent to the number of the segment on her score sheet.

In order to determine when the rater was ready to score real subject products, reliability checks were conducted during training. Three types of reliability were documented for this study: inter-rater, intra-rater, and subject. Inter-rater and Intra-rater were documented during rater training first, and then during the actual scoring process. Subject reliability was only documented during the actual scoring process. Inter-rater and Intra-rater reliability was documented by utilizing the percent agreement method (Herman et. al., 1992). A criterion score of .85 exact rater agreement was considered acceptable reliability. For inter-rater reliability, the rater's ratings of the 11 sample products were transferred from her individual score sheet to a master score sheet that already had the researcher's scores on it (Appendix I). The scores were then compared and a 91% percent reliability was determined between the researcher and the rater.

The rater also scored the same 11 sample products twice in a random order to determine if the rater's scores on products remained consistent throughout the scoring process (intra-rater). A criterion score of .85 exact rater agreement was considered acceptable reliability. For intra-rater reliability, the researcher documented the rater's scores on all 11 sample products twice and found the rater to be 100% reliable on her judgment of sample products (Appendix J).

The rater was more than .85% reliable with the researcher and was determined to be ready to score real products from this study. During the actual scoring process of real products, the rater would score subject products independently, and the researcher would check for inter-rater reliability as well as intra-rater reliability to obtain consistency in the scoring process.

Rater Training - Task Two

For Task Two, one rater and the researcher were used in order to create consistent, reliable scoring procedures on subject products. The rater was a doctoral candidate in physical education pedagogy and also had an extensive background in teaching physical education and tennis particularly at the collegiate level. It was evident from the background experience of the rater that she had the content knowledge concerning the basic skills of tennis. As in Task One, the rater and the researcher for Task Two completed the same course in Authentic Assessment prior to this study taking place. Therefore, the rater and the researcher had an understanding of authentic assessment, authentic assessment tasks, scoring rubrics, and reliability and validity issues.

To begin the rater training session, the rater received and reviewed a copy of Task Two. After reviewing the copy of the task, the rater had an opportunity to ask questions concerning the task. The following information was then discussed: the topic of the task, what the results could be used for, the subjects completing the task, the place where subjects would complete the task, the time allotted to subjects to complete the task, and the administration of the task.

General directions for Task Two were also reviewed so that the rater would have a clear understanding of what was expected of subjects. For example, subjects were to include the following items in their booklet: (1) the four basic skills in order for someone to play a game of tennis, (2) the correct grip for each skill, (3) three cues or tips for each skill to help children grasp the important information about the swing, (4) three appropriate ways to practice each skill, and (5) a cover for their booklet.

The rater was then given a copy of the scoring rubric for Task Two and was provided with a clear explanation of the scoring criteria. Criteria dimensions and scale values were defined and examples of scoring criteria were available for her to view. A chart listing acceptable skills, grips, critical components, and appropriate ways to practice each skill was available for the rater to use during the training process and actual scoring process (Appendix K). The chart was developed by the researcher and based on the USTA Schools Program Curriculum. Therefore, the skills, cues, and practice methods located on the chart would be appropriate to use along with the scoring rubric in scoring subject products. The rater was reminded that there could be more additional acceptable responses not found on the chart. In other words, subject responses might have similar meanings to those listed on the chart, but be identified by a different name. If the rater and the researcher agreed to add some appropriate items then they could do so during rater training.

Prior to scoring sample products, the rater and the researcher viewed 8 sample training products that were previously coded by the researcher. Each value on the scoring criteria was represented twice by the training products. The entire range of subject performance likely to be encountered in the actual scoring process was also represented. A clear

justification for each rating was explained and discussed by the rater and the researcher. Practice scoring during rater training provided the rater with ample opportunities to ask questions and to receive immediate feedback about her judgments. As the rater became more proficient with the scoring process, she had opportunities to score additional sample training booklets and discuss borderline booklets. The rater also came to understand that her job was to make a judgment based on the scoring criteria and not on her previous experience or her own inclinations.

The rater then scored 10 sample products that were previously scored by the researcher from several pilot studies conducted prior to this study. Each booklet from the pilot studies was labeled with a number on the back of the booklet. The rater had her own score sheet. The rater scored each page of the booklet one at a time and then referred to the scoring rubric to check the scoring criteria. She then reported her score in the box adjacent to the number of the booklet on her score sheet. (Appendix L). The rater was instructed not to write on the sample products because the researcher did not want the rater to be influenced by her rating when it came time to check for intra-rater reliability.

According to Herman, Aschbacher, and Winters (1992), rater training ends when there is an acceptable level of agreement/consistency among raters. For the purposes of this study three types of reliability were used: inter-rater, intra-rater, and subject. Inter-rater and intra-rater reliability checks were documented by using the percent agreement method (Herman et. al., 1992). A criterion score of .85 exact rater agreement was considered acceptable reliability. For Inter-rater reliability, the rater's ratings on 10 sample products were transferred from her individual score sheet to a master score sheet that already had the researcher's scores on it (Appendix M). Inter-rater was then checked by the researcher by comparing the scores between the rater and the researcher and a percent reliability was determined. The scores were then compared and a 100% reliability was determined between the researcher and the rater.

The rater also scored the same 10 sample products twice in a random order during rater training to determine if her scores on products remained consistent throughout the scoring process (intra-reliability). A criterion score of .85 exact rater agreement was considered acceptable reliability. For intra-rater reliability, the researcher documented the rater's scores on all 10 sample products twice and found the rater to be 100% reliable in judging sample products (Appendix N).

Since the rater was more than .85% reliable with the researcher, she was determined to be ready to score real products from this study. The rater would score subject products independently during the actual scoring process, and the researcher would check the inter-rater reliability as well as intra-rater reliability.

### Subject Reliability

In addition to inter-rater and intra-rater reliability, subject reliability was determined to ensure subject responses and scores remained consistent over time (Herman et. al., 1992). Subject reliability was checked by giving the same task to subjects to complete on two separate occasions one week apart. For example, upon completion of a task, subjects attempted the same task one week later. Subject scores on both attempts were recorded and then subject reliability was determined during the actual scoring process.

### **Determining Validity**

This study used two methods to determine validity of the alternative assessment tasks: The selection of subjects was a measure of construct validity, and a review process by a United States Tennis Association Professional was a measure of content validity (Thomas & Nelson, 1990).

#### **Construct Validity**

Subjects used for this study were EUVTP's, EUTP's, and IUTP's. The EUVTP's consisted of six tennis players on the schools varsity tennis team. EUTP's consisted of six experienced tennis players participating in a regular basic instruction physical education class. Those who ranked highest (20 or above) on a questionnaire determining tennis knowledge and experience were chosen for this study. IUTP's consisted of six students participating in a regular basic instruction physical education class. Those who ranked lowest (10 or below) on a questionnaire determining tennis knowledge and experience were chosen for this study. It was hypothesized that EUVTP's and EUTP's would score higher on the tasks than IUTP's. If this was true, then it would serve as one measure of validity (Thomas & Nelson, 1990).

#### **Content Validity**

The researcher contacted the United States Tennis Association (USTA) in White Plains, New York, and enlisted the counsel of the Director of the Schools Program Curriculum. During a review process by this USTA professional, he was asked to reflect on the content validity of the assessment tasks and answer the following questions:

(1) Does the content of the assessment reflect the content in the USTA Schools Program Curriculum? and (2) Are the scoring criteria consistent with the goals of the USTA Schools Program curriculum? If USTA raters answered "yes" to these questions, than it would serve as another measure of validity (Thomas & Nelson, 1990).

## **Procedures**

Two alternative assessment tasks based on the USTA Schools Program were completed by experienced university varsity tennis players, experienced university tennis players and inexperienced university tennis players. The following section describes in detail the data collection procedures for this study (Appendix O).

### Experienced University Varsity Tennis Players (EUVTP's)

Six individuals chosen for this study were administered the first of two tasks at the university tennis courts before team practice. They were allotted 45 minutes to complete the first task. EUVTP's consisted of two groups of three subjects and each group was assigned a number, 1 or 2. In Task One, subjects worked together in groups of three to create a video explaining and demonstrating three different basic tennis skills. Prior to making the video, each subject wrote four cues or tips on a posterboard that helped them in their explanation of the skill. When a group finished their posters, they were then ready to make a video. One video camera was used to film subjects, therefore one group of subjects was filmed at a time. The group of subjects not being filmed continued to work together in preparing for their video. They were not allowed to observe the other group of subjects making their video.

When both groups completed Task One, the videotape containing the explanations and demonstrations of basic tennis skills was labeled "Experienced university varsity tennis players, Task One - trial one: groups 1 & 2." Posters were also collected from each group of subjects, but they were not scored.

Subjects completed this same task again one week later to determine subject reliability. For example, if Task One was given on Tuesday, then the following Tuesday subjects would attempt Task One again in order to determine subject reliability.

The same six subjects were administered Task Two, two days after the completion of Task One. The task was given at the university tennis courts before team practice and subjects were allotted 45 minutes to complete the task. In Task Two, subjects worked individually to develop a booklet of basic tennis skills. A folder containing four skill pages was distributed to each subject. In this booklet subjects were asked to include four basic skills needed to begin to play a game of tennis, three cues or tips for each skill, and three ways to practice each skill. Subjects were also asked to design a cover for their booklet.

When subjects completed this task, they turned in their booklet. As each booklet was turned in, they were coded with a number placed on the back of the folder. Numbers 1 - 6 were used for EUVTP's. Subject's names and the number placed on the back of their booklet were recorded to ensure that the numbers correspond with the subjects. Subjects attempted this same task again one week later to determine subject reliability.

### Experienced University Tennis Players (EUTP's)

EUTP's followed the same procedures as EUVTP's except for the following differences. Six individuals chosen for this study were administered both tasks at the university's tennis courts or in the gym depending on weather conditions. Subjects attempted Task One during their regularly scheduled basic instruction physical education class.

The same six subjects were administered Task Two, two days after the completion of Task One. Subjects attempted Task Two during their regularly scheduled basic instruction physical education class. Each booklet was then coded with a number placed on the back of the folder. Subjects completed this same task again one week later to determine subject reliability. The same coding process took place.

### Inexperienced University Tennis Players (IUTP's)

IUTP's followed the same procedures as EUVTP's except for the following differences. Six individuals chosen for this study were administered both tasks at the university's tennis courts or in the gym depending on weather conditions. Since these subjects were chosen from a regular basic instruction physical education class they attempted both tasks during their regularly scheduled class time. IUTP's consisted of two groups of three and each group was assigned a group number, 5 or 6.

The same six subjects were administered Task Two, two days after the completion of Task One. Subjects attempted Task Two during their regularly scheduled basic instruction physical education class. The same coding process took place.

## **Data Analysis**

The following section describes the determination of reliability and validity. The following methods of reliability were analyzed for this study: inter-rater, intra-rater and subject reliability. As described earlier, raters were provided a method for recording subject scores on each alternative assessment task during rater training. Two types of validity were also analyzed for this study: construct validity and content validity.

### Inter-Rater Reliability

The raters' ratings on both tasks were recorded during the scoring process and were analyzed by calculating rater agreement. As previously mentioned, a criterion score of .85 exact rater agreement was considered acceptable reliability. For example, in this study, six videos were completed and scored by the rater and the researcher. Since a criterion score of .85 was considered acceptable reliability, then 5 out of 6 times the rater and the researcher needed to be in exact agreement regarding their scoring.

Task two required subjects to work independently to develop a booklet of basic tennis skills. Upon completion, a total of 18 booklets were collected

and scored by the rater and the researcher. Since a criterion score of .85 was considered acceptable reliability, then 15 out of 18 times the rater and the researcher needed to be in exact agreement regarding their scoring.

#### Intra-Rater Reliability

The same procedure used to analyze inter-rater reliability was used to analyze intra-rater reliability. The rater scored the same task twice during the scoring process and her ratings for the same task were recorded. A criterion score of .85 was considered acceptable intra-rater reliability.

#### Subject Reliability

The same procedure and criterion score used to analyze inter-rater reliability was used to analyze subject reliability. Subject reliability was checked by giving the same task to subjects one week later in the study. Subjects' scores on both attempts of each task were recorded during the actual scoring process. After scores were recorded for all subjects, then subject reliability was calculated.

#### Construct Validity

Construct validity (Thomas & Nelson, 1990) was determined by analyzing the results of subjects' products to determine whether the data supported the hypotheses for this study. The scores on both tasks should be higher for experienced tennis players and lower for inexperienced players. If the scores of both tasks were different, then construct validity could be inferred.

#### Content Validity

Content validity (Thomas & Nelson, 1990) was determined by analyzing the responses of a USTA professional. If he agreed that both tasks measured the content in the USTA Schools Program Curriculum, then the tasks could be said to have content validity (Appendix P).

### **Summary**

The purpose of this study was to develop two valid and reliable alternative assessment tasks based on the USTA Schools Program Curriculum. These assessments might then be used to determine the effectiveness of the USTA Schools Program Curriculum as taught in selected schools. The subjects selected were experienced university varsity tennis players, experienced university tennis players, and inexperienced university tennis players. The first task required students to work in groups of three to create a video explaining and demonstrating basic tennis skills. The second task required students to work individually to develop a booklet of basic tennis skills.

Scoring rubrics were developed based on the goals and objectives of the USTA Schools Program Curriculum. The rubrics contained descriptions that served as the basis for judging subject performance. Subject products were collected and scored by trained raters.

Three types of reliability (intra-rater, inter-rater, and subject) were analyzed by calculating rater agreement. A criterion score of .85 was considered

acceptable reliability. Two types of validity (construct and content) were also analyzed. The scores on both tasks should be higher for experienced tennis players and lower for inexperienced tennis players. If the scores were different, then construct validity could be inferred. Content validity was determined by analyzing the responses of a USTA Professional. If he agreed that both tasks measured the content in the USTA Schools Program Curriculum, then the tasks could be said to have content validity.

## CHAPTER IV

### Results and Discussion

The purpose of this study was to develop two valid and reliable alternative assessment tasks based on the USTA Schools Program Curriculum. These assessments might then be used to determine the effectiveness of the USTA Schools Program Curriculum as taught in selected schools.

Task One required subjects to work in groups of three to create a video explaining and demonstrating basic tennis skills. Each subject chose a different basic tennis skill and explained the skill on video. Subjects also demonstrated the skill or asked someone else in their group to demonstrate the skill while they explained.

Task Two required subjects to work independently to develop a booklet of basic tennis skills. Subjects were to include the following items in order to develop their booklet: the four basic skills of tennis, the correct grip for each skill, three critical components for each skill, and three appropriate ways to practice each skill. This chapter will report the results produced from this study and will explain the meaning of the results.

#### **Results**

As described in Chapter Three, three types of reliability were established for Task One and Task Two. A criterion score of .85 exact rater agreement was considered acceptable reliability. Two types of validity were also determined for both alternative assessment tasks. The results are divided into five categories:

- 1) Inter-rater reliability
- 2) Intra-rater reliability
- 3) Subject reliability
- 4) Construct validity
- 5) Content validity

#### **Inter-rater Reliability**

Inter-rater reliability was the degree to which different raters obtained the same scores on the same product (Thomas & Nelson, 1990). The rater and the researcher's ratings on Task One and Task Two were recorded during the actual scoring process. A criterion score of .85 exact rater agreement was considered acceptable reliability for Task One and Task Two. For Task One and Task Two 100% inter-rater agreement was obtained.

#### **Intra-rater Reliability**

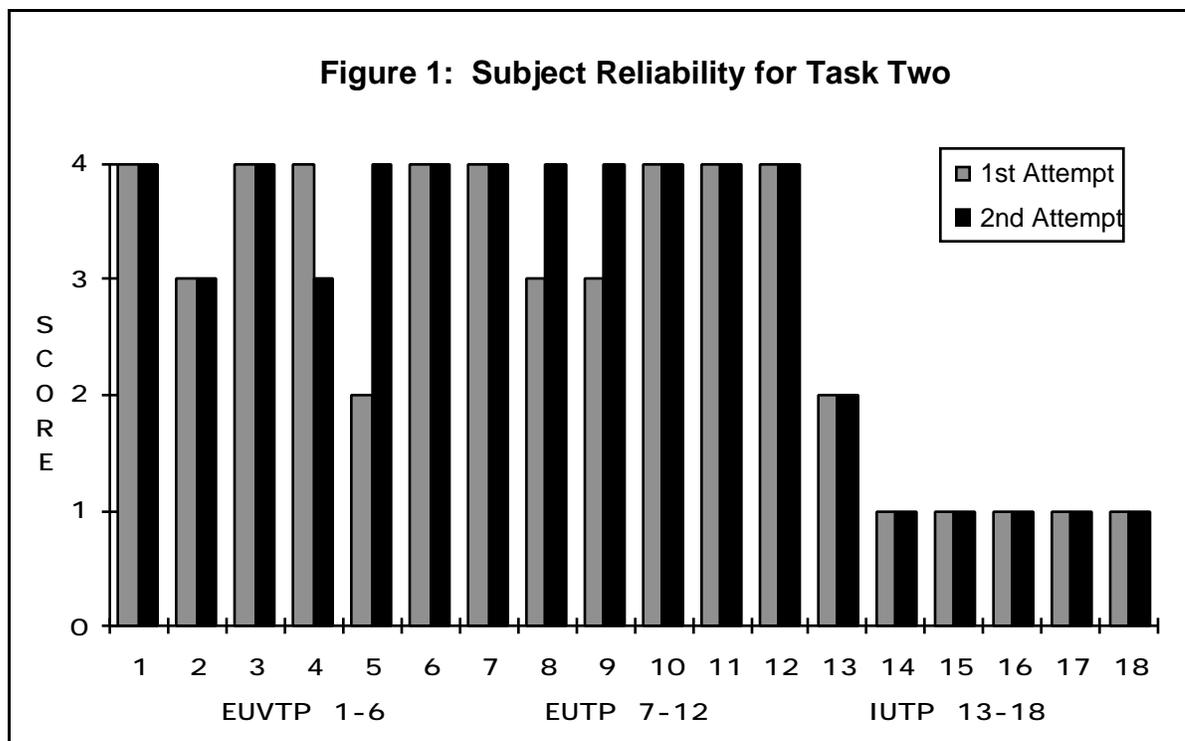
Intra-rater reliability was the degree to which one rater remained consistent over time (Herman et. al., 1992). The rater scored the same product twice during the actual scoring process. Subjects' scores were recorded by the rater, and the researcher checked intra-reliability. A criterion score of .85 exact rater agreement was considered acceptable reliability for both

assessments. For Task One and Task Two, 100% intra-rater reliability was obtained during the actual scoring process.

### **Subject Reliability**

Subject reliability was the degree to which subjects remained consistent over time (Herman et. al., 1992). Subject reliability was checked by giving Task One and Task Two to subjects during two trials in the study that were spaced one week apart. Subjects' products were scored by trained raters, then reliability between the first and second trials was calculated by using .85 exact rater agreement. For Task One, subjects were 100% reliable on their second trial.

For Task Two, subjects were 78% reliable on their second trial. Figure One illustrates the difference in subject scores on their first and second attempt on Task Two. EUVTP's were identified as numbers 1-6, EUTP's were identified as numbers 7-12, and IUTP's were identified as numbers 13-18.



## Construct Validity

Construct validity was determined by using subjects with known differences in ability or understanding to complete both tasks. If scores on Task One and Task Two were different between experienced university tennis players and inexperienced tennis players, then construct validity could be inferred (Thomas & Nelson, 1990).

Figure 2 illustrates the profiles of scores for Task One among experienced university varsity tennis players (EUVTP), experienced university tennis players (EUTP), and inexperienced university tennis players (IUTP). EUVTP's and EUTP's scored 4's which meant they were characterized as having an advanced level of tennis knowledge. The two groups of IUTP scored a 1 on a scale of 4, which meant they were characterized as having a beginner's level of tennis knowledge. Since the scores on Task One were higher among experienced university tennis players than inexperienced tennis players, then construct validity could be inferred for Task One.

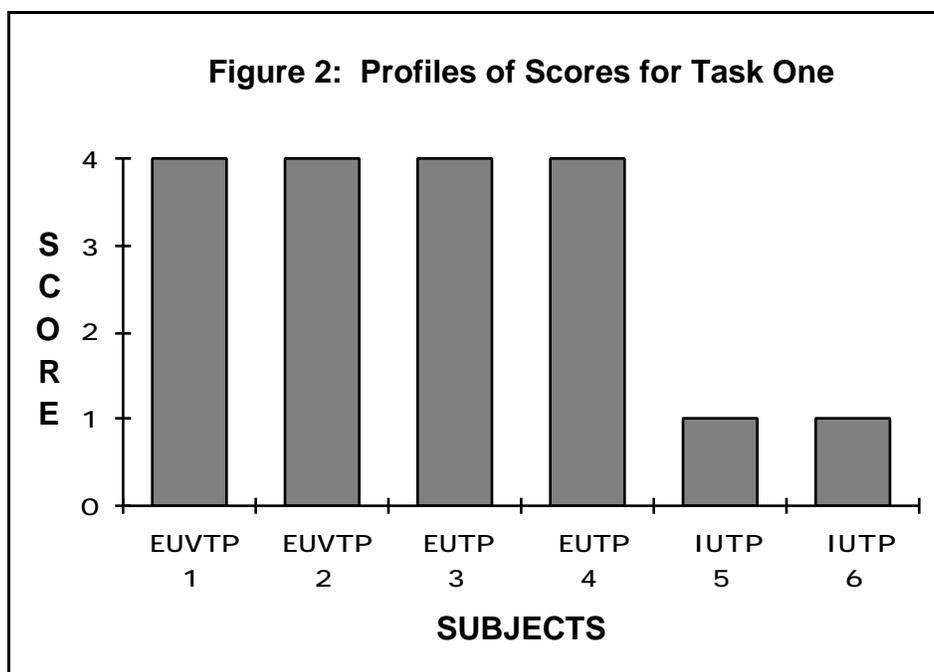


Figure 3 illustrates the profiles of subject scores for Task Two. EUVTP were assigned numbers 1-6, EUTP were assigned numbers 7-12, and IUTP were assigned numbers 13-18. EUVTP scores consisted of four 4's, one 3, and one 2. EUTP scores consisted of four 4's and two 3's. IUTP scores consisted of one 2, and five 1's. Since there was a difference in scores between experienced university tennis players and inexperienced tennis players, then construct validity could be inferred.

**“Insert Figure 3 here**  
**Figure 3, Profiles of Scores for Task Two**

**Content Validity**

Content validity was defined by asking the question "Does Task One and Task Two measure the content in the USTA Schools Program Curriculum?" (Thomas & Nelson, 1990). The director of the USTA Schools Program Curriculum in White Plains, New York, was asked to review Task One and Task Two and respond to the following statements by answering "strongly agree, agree, disagree, or strongly disagree." The director was also asked to comment on his response:

First, did the content of the assessment task reflect the content of the USTA Schools Program Curriculum, and second, was the scoring rubric consistent with the goals and objectives of the USTA Schools Program Curriculum.

For Task One, creating a video of basic tennis skills, the USTA Schools Program Director "strongly agreed" that the content of the assessment task reflected the content in the USTA Schools Program Curriculum. The program director "agreed" that the scoring rubric for Task One was consistent with the goals and objectives of the USTA Schools Program Curriculum. He commented, "The basic strokes of tennis are the forehand groundstroke, backhand groundstroke, serve, forehand volley and backhand volley. These skills are taught in the USTA Schools Program Curriculum, and therefore would be the only skills accepted for the video."

For Task Two, developing a booklet of basic tennis skills, the USTA Schools Program director "strongly agreed" that the content of the assessment task reflected the content in the USTA Schools Program

Curriculum. He commented "The basic skills of tennis, are the forehand groundstroke, backhand groundstroke, serve, forehand volley, and backhand volley. All of these skills are taught in the USTA Schools Program Curriculum. Some subjects could give different grip variations for a skill, and as long as the grip was appropriate for the skill, it would be considered an acceptable answer." The program director "agreed" that the scoring rubric for Task Two was consistent with the goals and objectives of the USTA Schools Program Curriculum. The program director commented "These Tasks will definitely show tennis knowledge, but I consider the technique description, stroke name and teaching cues more important than the appropriate ways to practice." Since the Program Director of the USTA Schools Program Curriculum "strongly agreed" and "agreed" with these statements for Task One and Task Two, then content validity can be inferred.

### **Discussion**

The purpose of this study was to develop two valid and reliable alternative assessment tasks based on the USTA Schools Program Curriculum. The following section will discuss the results and will present conclusions as they pertain to this study. The discussion is organized into three categories: 1) Inter-rater Reliability 2) Intra-rater Reliability 3) Researcher's Observations

#### **Inter-rater Reliability**

Inter-rater reliability was a measure of how closely two or more raters assign similar ratings to the same task (Thomas & Nelson, 1990). The purpose of inter-rater reliability was to make sure that ratings assigned were not influenced by the rater's own values, attitudes and other personality characteristics.

Based on the raters' ratings for this study, 100% inter-rater reliability was obtained for Task One and Task Two. 100% inter-rater reliability meant that the rater and the researcher gave the same score on the same products for Task One and Task Two, therefore obtaining consistency in scoring subject products. Research has shown that acceptable levels of rater agreement were obtained when the judges were (1) drawn from the same discipline, (2) used explicit scoring criteria, and (3) participated in a rater training session (Herman et. al., 1993; Deidrich, 1963). The following paragraphs provide a detailed explanation of how these areas influenced the reliability of Task One and Two.

The first reason 100% inter-reliability was obtained for Task One and Task Two, was probably because the raters and researcher were from the same discipline. The raters and the researcher were both physical education teachers who had previously taught tennis at the high school and university levels and were recreational players of the game. The researcher and the raters also enrolled in a university course entitled "authentic assessment" prior to launching this study. During this course, the researcher and the raters developed an understanding of authentic assessment tasks, scoring

criteria, and how rater training sessions can be used to develop consistency among rater's. Selecting raters with prior understanding of authentic assessment contributed to rater consistency in subject products.

The second probable reason 100% reliability was obtained for Task One and Task Two was the use of explicit scoring criteria. Prior to this study, the task and the scoring rubric were rewritten numerous times to ensure clarity not only for subjects completing the task but for rater's in scoring subject products. Two charts were developed by the researcher based on the USTA Schools Program Curriculum which listed the four basic tennis skills and critical components for Task One as well as the correct grip and appropriate ways to practice for Task Two. These charts made the scoring of subject products easier for the raters because they could refer to it if there was any doubt about how to score a subject's performance.

A final reason that 100% inter-reliability was obtained for Task One was the use of an extensive rater training program. The raters were oriented to the assessment task by receiving an overview of the assessment context, what the results could be used for, what directions and prompts the subjects would receive, and how the task was to be administered. They also received general directions for Task One so that they would have a clear understanding of what was expected of subjects.

The researcher also clarified the scoring criteria for the raters. The criteria dimensions and scale values were defined and examples of scoring criteria were available for the raters to view. As described above, charts were available for the raters to and some figures illustrating proper stroke analysis. The raters could refer to the chart if they had any questions pertaining to a skill.

The raters had an opportunity to view a range of training products with the researcher prior to scoring sample products on their own. Each product represented not only each value on the scoring criteria, but also the entire range of subject performance likely to be encountered in the actual scoring process. By doing this, the researcher gave the raters an opportunity to discuss the ratings in depth and provide a clear justification of each. The raters also came to understand that their job was to make a judgment based on the scoring criteria and not their own inclinations. After viewing the training products, the raters were asked to score sample products one at a time, were given a method for reporting scores. After they scored the sample products and were at least .85 in exact agreement with the researcher, then they were permitted to score real products. Providing the raters with a training session ensured that the raters applied the scoring criteria consistently to the range of subject products for this study.

### **Intra-rater Reliability**

Intra-rater reliability was the degree to which one rater remained consistent over time (Herman et. al., 1992). Raters scored all products twice in

random order during the actual scoring process. The scores obtained by the raters were then rated by the researcher. The ratings for both Task One and Task Two showed 100% intra-rater reliability. This level of intra-rater reliability for Task One and Task Two indicates that the raters were consistent in scoring subject products throughout the scoring process. The following paragraphs will discuss why 100% intra-reliability was obtained for this study.

The first probable reason why 100% intra-rater reliability was obtained was that the raters for Task One and Task Two were selected from the same discipline as the researcher. As discussed previously, the raters chosen to score subject products for this study had previous tennis knowledge obtained from teaching high school and university students as well as playing tennis for recreation.

The second probable reason why 100% intra-rater reliability was obtained was that raters for Task One and Task Two were trained in the morning on two different occasions. Training was scheduled in the morning, rather than in the afternoon because in general, people tend to tire and rate more slowly later in the day (Herman et. al., 1993). As a result of working when they were most alert, the raters were able to score subject products more consistently.

The third reason 100% intra-rater reliability was obtained was the explicit scoring criteria used. Task One and Task Two's rubrics were field-tested to assure they measured the subject product qualities for which they were designed. The criteria was clear and concrete, which made it easier for the raters to understand. Annotated examples of all score points were available for the raters to analyze and to use to practice scoring. Ample practice and feedback were available to the raters while scoring training products.

Finally, by training the raters to score subject products based on the scoring criteria rather than on the rater's own judgments or past experiences it was possible to obtain 100% intra-rater reliability.

### **Researcher's Observations**

It was interesting to observe the members of each subject group interact as they planned how they would accomplish each task. It appeared that the experienced subjects selected the basic skills they were going to demonstrate more quickly than the inexperienced subjects. The experienced subjects were more serious and put forth more effort in preparing for their performance on video and in developing their booklet. After completion of the alternative assessment tasks, subjects were asked to share their impressions of the alternative assessment tasks and reflect on their physical education assessment experiences from elementary through high school. Three of the inexperienced subjects reported that "they didn't enjoy physical education in school because when written tests

were given they passed them, and did poorly on skill tests. They would end up feeling totally frustrated because they felt they were never taught how to become proficient at a skill nor were they taught the "critical components" of a skill." Two of the inexperienced subjects didn't have tennis as part of their physical education experience and one of the inexperienced subjects could only recall being tested on the Presidential Physical Fitness test in school.

The experienced subjects related that passing a physical education tennis unit in school was a "breeze" because their ability and experience helped them to ace the skill tests when given. However, they reported that the alternative assessments in this study were significantly more challenging than the skill or written tests they were accustomed to taking. They felt the alternative assessments were a more thorough test of knowledge. The assessment tasks challenged them to think through the steps necessary to perform the skill instead of merely performing it based on instinct and repetition.

### **Summary**

Two valid and reliable alternative assessment tasks were developed in this study based on the USTA Schools Program Curriculum. Task One required subjects to work in groups of three to create a video explaining and demonstrating basic tennis skills. Task Two required subjects to work independently to develop a booklet of basic tennis skills. Three types of reliability were established for both tasks: inter-rater, intra-rater, and subject. Two types of validity were also determined for both alternative assessment tasks: construct validity and content validity.

Inter-rater reliability was established at 100% for the following reasons: the raters and the researcher were from the same discipline, explicit scoring criteria was used, and rater's were extensively trained. Intra-rater reliability was also established at 100 % for similar reasons: the raters were selected from the same discipline as the researcher, explicit scoring criteria was used, and rater's were trained to score based on set criteria rather than their own judgements or past experiences. Subject reliability differed between the two tasks, with Task One subjects being 100% reliable from trial one to trial two, and Task Two subjects being 70% reliable on their second trial. The differences in reliability on Task Two were attributed to differences in subject motivation and effort as well as the possibility of additional learning that might have taken place between the two trials. Construct validity was obtained based on a difference in task scores between tennis players with differing experience levels in the game. Content validity was obtained based on professional analysis which agreed that both tasks measured the content in the USTA Schools Program Curriculum.

Inexperienced subjects selected for this study reflected on their physical educational experiences and reported the following: they passed written tests when given but did poorly on standardized skill tests, did not learn the

"critical components" of skills, and one did not have tennis as part of his physical education experience.

Experienced subjects selected for this study reflected on their physical educational experiences and reported that the alternative tests developed for this study were a more thorough tests of their cognitive knowledge as compared to the tests they were given in their physical educational experiences.

## CHAPTER V

### Summary and Conclusions

This chapter will be divided into the following sections:

- 1) Review of the Study
- 2) Summary of Conclusions
- 3) Significance of Research
- 4) Applications of Research to Physical Education Programs
- 5) Applications of Findings
- 6) Recommendations for Further Research

#### Review of Study

The purpose of this study was to develop two valid and reliable alternative assessment tasks based on the goals and objectives of the USTA Schools Program Curriculum. These assessments might then be used to determine the effectiveness of the USTA School's Program Curriculum as taught in selected schools.

The two alternative assessment tasks developed for this study were based on the goals and objectives of the USTA Schools Program Curriculum. The first task required subjects to work in groups of three to create a video explaining and demonstrating basic tennis skills. The second task required subjects to work individually to develop a booklet of basic tennis skills. Scoring rubrics were also developed based on the goals and objectives of the USTA Schools Program Curriculum. The rubrics contained descriptions that would serve as the basis for judging subject products.

Subject products were collected, coded, and then scored by a trained rater. The rater was trained extensively in order to retain reliability. Three types of reliability methods (intra-rater, inter-rater, and subject) were used to determine the reliability of the assessment tasks, and they were analyzed by calculating rater agreement. A criterion score of .85 exact rater agreement was considered acceptable reliability.

Two types of validity were determined as well. Construct validity was determined by analyzing the results of subjects' products to determine whether the data supported the hypothesis for this study. Content validity was determined by analyzing the response of the USTA professionals concerning the content of the assessment tasks and rubrics.

Two research questions were developed for this study: Was it possible to develop satisfactory levels of validity on Task One and Task Two? and Was it possible to develop satisfactory levels of reliability on Task One and Task Two based on the USTA Schools Program Curriculum? Based on the results of this study as described in Chapter Four, acceptable levels of validity and reliability were obtained for Task One: Creating a Video of Basic Tennis Skills and Task Two: Developing a Booklet of Basic

Tennis Skills. Task One and Task Two could therefore be used as another means to assess what students have learned as a result of participating in a USTA Schools Program Curriculum.

### **Summary of Conclusions**

There were four hypotheses examined in this study, three pertaining to subjects and one pertaining to the assessment tasks. The three subject related hypothesis were: EUVTP's would score better on Task One and Task Two as compared to EUTP's and IUTP's; EUTP's would score better on Task One and Task Two than IUTP's; and IUTP's would score lower on Task One and Task Two as compared to EUVTP's and EUTP's. It was also hypothesized that it is possible to develop valid and reliable alternative assessment tasks based on the USTA Schools Program Curriculum. The results of the study proved the first hypothesis to be false, and the remaining three to be true.

The researcher concludes that the first hypothesis was false because the subjects in the EUTP group had a knowledge base of tennis equivalent to the EUVTP group even though they were not on the university varsity tennis team. Since the assessment tasks primarily evaluated cognitive knowledge of tennis rather than skill levels, it becomes apparent that any subjects with equal knowledge will score similarly regardless of how well they can actually play the game.

It can be concluded that the second and third subject related hypothesis were true because the subjects in the EUTP group had significantly more prior knowledge of tennis than the IUTP group based on their questionnaire scores and their results of Task One and Task Two. A final conclusion reached was that the assessment tasks hypothesis was true because the tasks were designed in such a way that they measured what they were intended to measure, basic tennis knowledge.

### **Significance of Research**

If physical education wants to assume the level of importance in school curriculums that the professionals in the field believe is lacking, then change in assessment is essential. One of the criticisms of traditional assessments is that they do not match program goals or hold students accountable for instructional goals. Since traditional assessments have not been able to measure what students have actually learned, the credibility of physical education teachers and programs has been weakened. Alternative assessments provide meaningful measures of "students' progress toward real-life outcomes that can't always be measured by standardized tests." (Hopple, 1995, p. 20). It is the researcher's opinion that the accountability and credibility of physical education teachers and programs would improve significantly if alternative assessments were implemented.

Physical education teachers need to determine the exit level outcomes of their programs as well as the program, course, unit, and lesson outcomes prior to beginning instruction (Hopple, 1995). Alternative assessment tasks

need to be developed to match the outcomes of the program and then implemented to improve learning and instruction. Once this has been accomplished teachers can hold students accountable for reaching their instructional goals. The products produced by students in alternative assessment tasks can also be used by physical education teachers in times of program evaluation to justify their program's existence.

Alternative assessment tasks can be beneficial to teaching and learning in physical education in other ways as well. They can provide feedback to teachers which can be used to improve individual student learning, class instructional methods, and entire physical education programs. These tasks help teachers base their instruction on the "real, not presumed, needs" of their students (Hopple, 1995, p. 21). The results of these assessments can be used to demonstrate to administrators, parents, and the community what students are learning in physical education (Kimball, 1996). Other benefits include students using creativity to explain their own knowledge of a motor skill and learning to use a variety of resources such as the Internet, newspapers, or videos to help them complete and assignment (Jefferies et. al., 1996).

### **Application of Findings**

The findings in this study can be applied to both physical education program development and program justification.

#### **Application of Findings to Physical Education Programs**

Alternative assessments take time to design, implement, and evaluate. However, there are ways in which physical education teachers could utilize alternative assessments to enhance their programs without being intimidated or overloading their already hectic schedules. Alternative assessments can be used: (1) when students leave a program to evaluate and provide justification for the program, (2) as ongoing assessments of students' progress e.g., pulling aside a few students at a time to participate in alternative types of assessment while the rest of the class remains on task, (3) in cooperation with teachers in other subject areas to avoid duplication of work, e.g. writing assessment to evaluate knowledge in physical education may also be used to evaluate writing skills in the language arts program, (4) as a base for networking with physical education with physical education colleagues, (5) to set realistic goals for developing and implementing new assessments each year addressing different aspects of a program (Mustain, 1994).

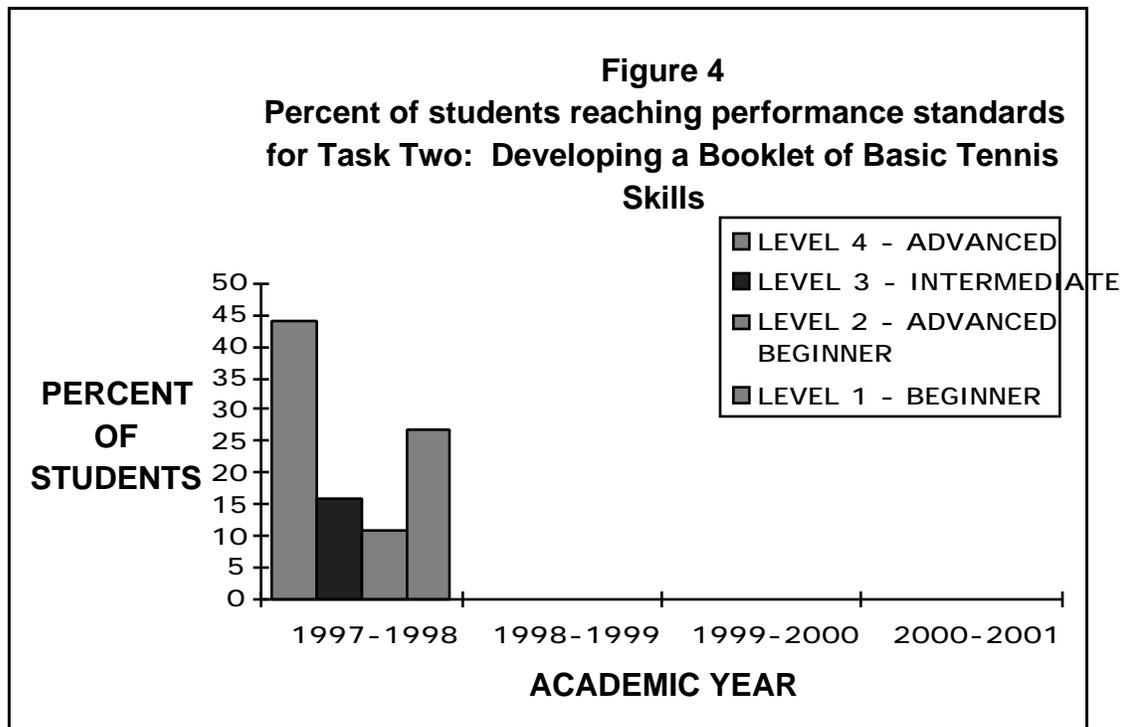
Physical education teachers should decide based on their teaching situations, how and when to use the alternative assessments. Scoring rubrics need to be designed to match the program goals and be able to differentiate between what is a quality performance and what is not.

#### **Application of Findings to Physical Education Program Justification**

Reporting the results of an assessment can provide feedback about teaching methods and student progress. However, numbers alone only tell

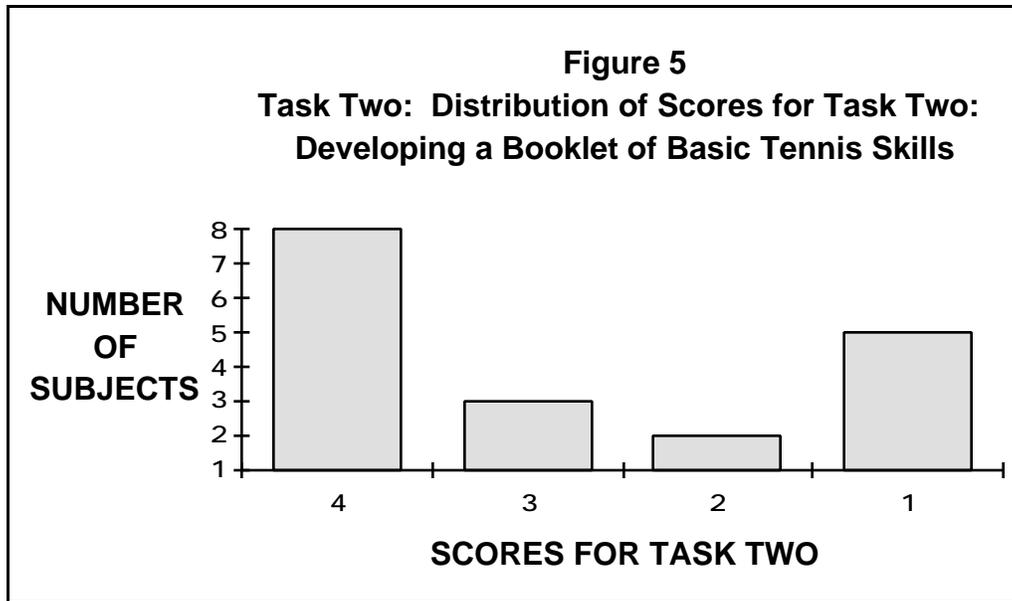
part of the story. Actual student performance products can provide the additional information needed to form a complete picture of what the program has accomplished.

There are several ways teachers can present results of their assessments to parents and administrators. One basic way is to show results is by graphing the percents of students reaching performance standards (Herman et. al., 1992). The following figure is an example of this using the results for Task Two in this study.



Data can be compiled over a period of years to show trends in student achievement and allow teachers to more concretely justify their programs.

Another way to present assessment results is to look at the range of student performance on a task and calculate how many students received each possible score (Herman et. al., 1992). This would be particularly helpful when there was no preconceived notion of how students would perform. The following figure provides an example of a distribution of student scores using the results for Task Two in this study.



### Recommendations for Further Research

This study focused on alternative assessment that might be used as a means of evaluating student learning in the USTA Schools Program Curriculum. The researcher recommends that this study be replicated using a subject group selected from the population of schools which have utilized the USTA Schools Program Curriculum. The format could be structured so that the assessment tasks are given as pre and post tests.

The researcher would also like to see more research on alternative assessment in physical education, with emphasis on: (1) authentic forms of alternative assessment, (2) development of undergraduate courses in alternative assessment, (3) the use of alternative assessment in the student teaching experience, (4) the use of alternative assessment in "rookie" year or beginning teacher programs prior to receipt of state teaching certification, and (5) developing more alternative assessment tasks specifically designed for physical education programs.

The body of research literature for alternative assessment in academics is extensive. It is critical for physical education professionals to bring their own volume of research in this area to equal level. Doing so will arm them with powerful instrument for change and "create an atmosphere for learning in physical education that can excite our students and rekindle a national interest in daily physical education for .... students." (Veal, 1992, p.91).

## REFERENCES

- Anderson, W. (1988). Preparing and Using the Written Curriculum. Journal of Physical Education, Recreation, and Dance, 67-72.
- Avery, C., Richardson, P., & Jackson, A., (1979). A Practical Tennis Serve Test: Measurement of Skill Under Simulated Game Conditions. Research Quarterly, 50 (4), 554-564.
- Brandt, R. (1992). On Performance Assessment: A Conversation with Grant Wiggins. Educational Leadership, 49 (8), 35-37.
- Brandt, R. (1993). On Teaching for Understanding: A Conversation with Howard Gardner. Educational Leadership, 50 (7), 4-7.
- Bruder, I. (1993). Alternative Assessment: Putting Technology to the Test. Electronic Learning, 12 (4), 22-29.
- Buschner, C. (1994). Teaching Children Movement Concepts and Skills: Becoming a Master Teacher. Champaign, IL: Human Kinetics.
- Cobane, E. (1962). Test for the Service. Tennis-Badminton Guide Washington D.C.: AAHPER.
- Cronin, J. (1993). Four Misconceptions about Authentic Learning. Educational Leadership, 50 (7), 78-80.
- Deidrich, P.B. (1963). "The Measurement of Skill in Writing" School Review 54, 584-592.
- DiGennaro, J. (1969). Construction of Forehand Drive, Backhand Drive, and Service Tennis Tests. Research Quarterly, 40, 496-501.
- Driver, H. I. (1973). Tennis for Teachers. Madison, WS: Monona-Driver Books.
- Freeland, K. & Moore, J. (1996). The Six Assumptions of Assessment. Teaching Elementary Physical Education, 7 (2), 22-23, 26.
- Graham, G. M., (1992). Teaching Children Physical Education: Becoming a Master Teacher. Champaign, IL: Human Kinetics
- Graham, G. M., & Sheffield, J. (1993). United States Tennis Association Schools Program Curriculum. New York: United States Tennis Association.
- Hensley, L. (1997). Alternative Assessment for Physical Education. Journal of Physical Education, Recreation, and Dance, 68 (7), 19-24.
- Hensley, L., Lambert, L., Baumgartner, T., & Stillwell, J. (1987). Is Evaluation Worth the Effort? Journal of Physical Education Recreation and Dance, 58 (6), 59-62.
- Herman, J. (1992). What Research Tells Us About Good Assessment. Educational Leadership, 8, 74-78.
- Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). A Practical Guide to Alternative Assessment. Virginia: Association for Supervision and Curriculum Development.
- Hewitt, J.E. (1966). Hewitt's Tennis Achievement Test. Research Quarterly, 37, 231-240.

- Hopple, C. (1994). A Qualitative Analysis: What Less-Fit and More-Fit Fourth and Fifth Graders Think, Feel, and Know about Physical Fitness Tests. Unpublished master's thesis. Virginia Polytechnic Institute and State University. Blacksburg, VA.
- Hopple, C. (1995). Teaching for Outcomes in Elementary Physical Education. Champaign, IL: Human Kinetics
- Hopple, C. (1997). The Real World Process of Assessment. Teaching Elementary Physical Education, 8 (4), 4-7.
- Imwold, C., Rider, R., & Johnson, D. (1982). The Use of Evaluation in Public School Physical Education Programs. Journal of Teaching in Physical Education, 2 (1), 13-18.
- Jefferies, S., Jefferies, T., & Mustain, W. "Why Assess in PE?" PE Central. 16 Apr. 1997. Online.
- Joyner, A., & McManis, B. (1997). Quality Control in Alternative Assessment. Journal of Physical Education, Recreation, and Dance, 68 (7), 38-40.
- Kelly, L. (1989). Instructional Time: The Overlooked Factor in Physical Education Curriculum Development. Journal of Physical Education, Recreation, and Dance, 60 (5), 29-32.
- Kimball, R. (1996). Collaborating on Assessment. Teaching Elementary Physical Education, 7 (2), 13.
- Kraft, E. F. (1979). Preparing for Tennis in the 1980s. Jopher, 39.
- Lund, J. (1992). Assessment and Accountability in Secondary Physical Education. Quest, 44, 352-360.
- Lund, J. (1997). Authentic Assessment: It's Development and Applications. Journal of Physical Education, Recreation, and Dance, 68 (7), 25-28, 40.
- Madaus, G., & Kellaghan, T. (1993). The British Experience with 'Authentic' Testing. Phi Delta Kappan, 74 (6), 458-469.
- Marzano, R., & Costa, A. (1988). Question: Do Standardized Tests Measure General Cognitive Skills? Answer: No. Educational Leadership, 45 (8), 66-71.
- Melograno, V. (1997). Integrating Assessment Into Physical Education Teaching. Journal of Physical Education, Recreation and Dance, 68 (7), 34-37.
- Mitchell, R. (1992). Testing for Learning: How new approaches to evaluation can improve American Schools. New York: The Free Press.
- Mustain, W. (1994) Authentic Assessment: More Than Grist for the Grading Mill. Teaching Elementary Physical Education. 5 (3), 1, 4-5.
- National Tennis Foundation Inc. (1974). Official Rating Handbook. New Monmouth, N.J.: Talent-N-Timing/Tennis.
- Newmann, F. M. , & Wehlage, G. G., (1993). Five Standards of Authentic Instruction. Educational Leadership, 50 (7), 8-12.
- O'Neil, J. (1992). National System of Standards, Exams Piloted. Association for Supervision and Curriculum Development Update, 34 (8), 1-5.
- Pool, C. & Bracey, G. W. (1992) Making Sense of Authentic Assessment.

- Instructor, 40-41.
- Publication manual of the American Psychological Association (3rd ed.). (1983). Washington, DC: American Psychological Association.
- Schiemer, S. (1996). A Positive Learning Experience. Teaching Elementary Physical Education, 7 (2), 4-6.
- Schnitzer, S. (1993). Designing an Authentic Assessment. Educational Leadership, 50, (7), 32-35.
- Shepard, L. (1989). Why We Need Better Assessments. Educational Leadership, 46 (7), 4-9.
- Sternberg, R. (1989). The Tyranny of Testing. Learning, 60-63.
- Stiggins, R. (1991). Facing the Challenges of a New Era of Educational Assessment. Applied Measurement in Education, 4 (4), 263-273.
- Strand, B. N., & Wilson, R. (1993). Assessing Sport Skills. Champaign, IL: Human Kinetics
- Symons, C.W. & Gascoigne, J.I. (1990). The Nation's Health Objectives -- A Means to School-Wide Fitness Advocacy. Journal of Physical Education, Recreation and Dance, 61 (6), 59-63.
- Thomas, J.G., & Nelson, J.K. (1990). Research Methods in Physical Activity. Champaign, IL: Human Kinetics
- Veal, M. L. (1992). The Role of Assessment in Secondary Physical Education - A Pedagogical View. Joperd, 88-92.
- Wiggins, G. (1987). Creating a Thought-Provoking Curriculum. American Educator, 10-17.
- Wiggins, G. (1989a). Teaching to the (Authentic) Test. Educational Leadership, 46 (7), 233-237.
- Wiggins, G. (1989b). A True Test: Toward More Authentic and Equitable Assessment. Phi Delta Kappan, 70 (9), 703-713.
- Wiggins, G. (1992). Creating Tests Worth Taking. Educational Leadership, 26-33.
- Wood, T. (1996). We Have the Standards. Now What? Teaching Elementary Physical Education, 7 (2), 20-21.
- Zhu, W. (1997). Alternative Assessment: What, Why, How. Journal of Teaching Physical Education, Recreation, and Dance, 68 (7), 17-18.

## APPENDIX A

### Informed Consent

#### Division of Health and Physical Education Virginia Polytechnic Institute and State University

##### Informed Consent

I, \_\_\_\_\_, do hereby voluntarily agree and consent to participate in a research project conducted by Karyn Mullholand of the Division of Health and Physical Education of Virginia Polytechnic Institute and State University.

**Title of Study:** Alternative Assessment in Tennis.

**The purposes of this experiment include:**

1. To develop two valid and reliable alternative assessment tasks based on the USTA School's Program Curriculum. These tasks can then be used to determine the effectiveness of the USTA Schools Program Curriculum as taught in schools.
2. To develop alternative assessment procedures to be used in physical education.

**I voluntarily agree to participate in this research study. It is my understanding that my participation will include:**

The completion of two alternative assessment tasks based on the USTA Schools Program Curriculum.

**I understand that participation in this experiment may produce certain discomforts and risks. These discomforts and risks include:**

Not Applicable.....

**I understand that certain personal benefits may also be expected from participation in this experiment. These include:**

Not Applicable.....

**I understand that there may be one or more appropriate alternative procedures that might be advantageous to me.**

**These include:**

Not Applicable.....

I understand that any data and/or videotapes will be held confidential and may be used in future classes and/or presentations at professional meetings.

I understand that I may abstain from participation in any part of the experiment or withdraw from the experiment should I feel the activities might be injurious to my health. The experimenter may also terminate my participation should he feel that the activities might be injurious to my health.

I understand that it is my personal responsibility to advise the researchers of any preexisting medical problem that may affect my participation or of any medical problems that might arise in the course of this experiment and that no medical treatment or compensation is available if injury is suffered as a result of this research. A telephone is available which would be used to call the local hospital for emergency service.

I have read the above statements and have had the opportunity to ask questions. I understand that the researchers will, at any time answer my inquiries concerning the procedures used in this experiment.

Date\_\_\_\_\_ Time\_\_\_\_\_ am/pm

Participant signature\_\_\_\_\_

Witness\_\_\_\_\_

Project Director:\_\_\_\_\_ Telephone\_\_\_\_\_

HPE Human Subjects Chairman: Dr. Richard Stratton

Telephone: 231-5617

Dr. Ernie Stout, 301 Burruss Hall, 231-5281

To receive the results of this investigation, please indicate this choice by marking in the appropriate space provided below. A copy will then be distributed to you as soon as the results are made available by the investigator. Thank you for making this important contribution.

\_\_\_\_\_ I request a copy of the results of this study.

## APPENDIX B

### Tennis Questionnaire

**1. How often do you play tennis?**

daily (4) once a week (3) once a month (2) once a year (1) other (0)  
If other, please describe.....

**2. Have you ever taken tennis lessons?**      yes(4) no(0)

If yes, please describe

**3. Did you play on your high school tennis team?**yes(4) no(0)

**4. Have you played competitively in any USTA**yes(4) no(0)

**sponsored tournaments or any other sponsored  
tournaments? If so, please list them.**

**5. Do you know the rules related to serving in tennis?**

yes(4)      no(0)

**6. Do you know how to keep score in tennis?**

yes(4)      no(0)

**7. Did you participate in a tennis class in the basic  
instruction physical education program?**      yes(4)

no(0)

**Points = \_\_\_\_\_**

**Total**

**APPENDIX C**  
**TASK ONE...**  
**CREATING A VIDEO OF BASIC TENNIS SKILLS**

**SITUATION:** Due to recent reductions in educational funding, states across the country have been forced to make cuts in physical education programs. As a result, many students are not exposed to certain lifetime activities. One such activity is tennis. I need your assistance in making a video explaining and demonstrating basic tennis skills. This video could be used to teach students who do not have tennis as part of their p.e. program the basic skills of tennis.

**TASK:** I would like for you to work with two other students to create a video explaining and demonstrating a basic tennis skill so that someone with no tennis experience would be able to learn how to begin to play a game of tennis. I will provide posterboard, markers, tennis balls/rackets, and any other equipment you may need to complete this task.

**Complete the following steps in order to make your videotape.**

**1.** Select two other people who you can work with to create a videotape of basic tennis skills. Each person in your group must select a different basic tennis skill. Working together in groups of three, each person will explain and demonstrate a different basic tennis skill for someone who has never played tennis. If you do not want to demonstrate the skill, you may ask others in your group to demonstrate while you explain. You may help each other in your group.

**2.** After each person has chosen a different basic tennis skill, each of you needs to write on a poster board to help with your explanations of the skill. Try to list at least four cues or tips that will help someone do that skill well. Make sure you print large and clearly so someone could read your poster on video.

**3.** You may start working on your posters now. When everyone in your group is finished making the posters, you may want to talk to others in your group and decide who will explain 1st, 2nd, and 3rd, and who will demonstrate the skill while you explain. (Remember, you are allowed to explain and demonstrate the skill you chose.) You may also want to decide who will hold the poster while you explain the skill.

**4.** Each person in your group will have four minutes to explain and demonstrate a basic tennis skill on video. Remember, you may use your poster to refer to when explaining how to perform the skill. You, or someone else in your group needs to include a complete demonstration to go along with the explanations of each skill. Tennis equipment will be available for you as well as any other equipment you may need.

## APPENDIX D

### TASK TWO.... DEVELOPING A BOOKLET OF BASIC TENNIS SKILLS

**SITUATION:** Many parents enroll their children at neighborhood recreation centers because they cannot afford to stay home and entertain them during summer vacation. Because the enrollment level is typically high at these centers, there is a need for high school students, who are looking for summer jobs, to assist and help teach the children certain sport skills, specifically tennis. The supervisor in charge of the neighborhood recreation centers is looking for some booklets describing the basic strokes of tennis. The supervisor feels that the content provided in these booklets could help high school students teach tennis to children at each of the neighborhood recreation centers.

**TASK:** I would like for you to develop a booklet of basic tennis skills that other high school students could use to teach children the basic skills of tennis. I will provide writing utensils, markers, and any other materials that you may need to complete your booklet. All you need to do is fill in as much information as you can. You have 45 minutes to develop this booklet.

**Complete the following steps in order to develop your booklet:**

1. Think about the basic skills necessary to begin to play a game of tennis.
2. Then, include the following items in order to develop your booklet. (The pages of the booklet are provided for you. There are four pages total.)

(A) The four basic skills in order for someone to play a game of tennis (one skill on each page).

(B) The correct grip for each skill.

(C) Three cues or tips for each skill. These cues or tips will help children grasp important information about the swing.

(D) Three ways to practice each skill.

(E) Be sure to design a cover for your booklet.

3. When you have finished developing your booklet, you may turn it in.

## APPENDIX E

### Task One: Scoring Rubric

#### Level 4: ADVANCED

Subjects make a videotape and at least four critical components for each of the three skills are clearly and correctly explained and demonstrated.

#### Level 3: INTERMEDIATE

Subjects make the videotape and at least three critical components for two of the three skills are clearly and correctly explained and demonstrated. The other one is either incomplete or not explained (demonstrated) at all or three critical components are clearly and correctly explained and demonstrated for all three skills.

#### Level 2: ADVANCED BEGINNER

Subjects make the videotape and a minimum of three critical components for one of the three skills are clearly and correctly explained and demonstrated or one critical component is clearly and correctly explained and demonstrated for all three skills.

#### Level 1: BEGINNER

Subjects choose not to make a video because they have no tennis experience or subjects make a video, but show no understanding of the critical components used to perform the skill and they are demonstrated incorrectly.

## **APPENDIX F**

### **Task Two: Scoring Rubric**

#### **LEVEL FOUR: ADVANCED**

Subject develops a booklet and each page (4 pages total) includes the following:

- \* The name of a basic tennis skill (forehand, backhand, serve, forehand volley, backhand volley)
- \* A correct name or description of the proper grip of the skill chosen
- \* Three correct critical components or cues for the skill
- \* Three appropriate ways to practice the skill

**OR**

- \* The booklet contains 1-2 incorrect items or 30-32 correct items

#### **LEVEL THREE: INTERMEDIATE**

Subject develops a booklet and three of the four pages include the following:

- \* The name of a basic tennis skill (forehand, backhand, serve, backhand volley, forehand volley)
- \* A correct name or description of the proper grip of the skill
- \* Three correct critical components or cues for each skill
- \* Three appropriate ways to practice the skill for each skill.

**OR**

- \* The booklet contains 3 - 8 incorrect items or 24-29 correct items

#### **LEVEL TWO: ADVANCED BEGINNER**

Subject develops a booklet and two of the four pages include the following:

- \* The name of a basic tennis stroke (forehand, backhand, serve, forehand volley, backhand volley)
- \* A correct name or description of the proper grip
- \* Three correct critical components or cues
- \* Three appropriate ways to practice the stroke or at least one appropriate way to practice the stroke.

**OR**

- \* The booklet contains 9 -16 incorrect items or 16-23 correct items

**LEVEL ONE: BEGINNER**

Subject chooses not to develop a booklet due to lack of tennis knowledge or subject develops a booklet and attempts to complete only some of the items listed on each skill page:

backhand, serve,  
forehand volley, backhand volley)

- \* The name of one basic tennis stroke (forehand,
- \* A correct name or description of the proper grip
- \* One or two correct critical components or cues

**OR**

15 correct items

- \* The booklet contains 17 or more incorrect items or 1-

## APPENDIX G

### CHART OF BASIC SKILLS & CRITICAL COMPONENTS FOR TASK ONE

#### TASK ONE

<b>Forehand Groundstroke</b>	<b>Backhand Groundstroke</b>	<b>Serve: Simplified, Beginners, Punch or Full-swing</b>
shake-hands grip eastern-forehand grip forehand grip "V" of hand lined up with racket edge ready position side to target firm wrist racket back to fence watch the ball contact at front foot swing low to high follow-through return to ready position	1/4 turn inward knuckle on top twist left eastern-backhand grip side to target watch the ball racket back to fence swing low to high step across and forward a foot in front of foot point to the sky follow-through return to ready position	shake-hands grip eastern forehand grip continental grip forehand grip "V" of hand lined up with racket edge. toss straight up side to target palm to ear toss to the tip eye on ball down together/up together reach for the sky make an "x" follow through across body
<b>Forehand Volley</b>	<b>Backhand Volley</b>	
shake-hands grip firm wrist eastern forehand grip forehand grip "V" of hand lined up with racket edge. thumb to belly button side to target watch the ball contact out in front short back-swing block/punch no follow-through split-step rackethead above handle	eastern-backhand grip twist left knuckle on top 1/4 turn left shakehands grip firm wrist side to target watch the ball split-step step across and forward racket head above handle contact out in front block/punch no follow-through	

USTA SCHOOLS PROGRAM CURRICULUM (1993)

## APPENDIX H

### RATER TRAINING: METHOD FOR REPORTING SCORES

TASK ONE:  
Sample Products from Pilot Study

<b>Task 1: Video</b>	<b>Rater 's Score</b>
1	4
2	1
3	4
4	2
5	2
6	3
7	2
8	4
9	3
10	1
11	2

**APPENDIX I**  
**RATER TRAINING: METHOD FOR REPORTING**  
**SCORES**

Rater Training - Inter-rater Reliability

<b>Task 1: Video</b>	<b>Rater</b>	<b>Researcher</b>	<b>% Reliability</b>
1	4	4	yes
2	1	1	yes
3	4	4	yes
4	2	2	yes
5	2	2	yes
6	3	3	yes
7	2	2	yes
8	4	4	yes
9	3	2	no
10	1	1	yes
11	2	2	yes
			91%

## APPENDIX J

### RATER TRAINING: METHOD FOR REPORTING SCORES

#### TASK ONE

#### Rater Training - Intra-rater Reliability

<b>Task 1: Video</b>	<b>Rater 1st trial</b>	<b>Rater 2nd trial</b>	<b>% Reliability</b>
1	4	4	yes
2	1	1	yes
3	4	4	yes
4	2	2	yes
5	2	2	yes
6	3	3	yes
7	2	2	yes
8	4	4	yes
9	3	3	yes
10	1	1	yes
11	2	2	yes
			100%

## APPENDIX K

### CHART OF BASIC SKILLS, GRIPS, CRITICAL COMPONENTS, &

#### WAYS TO PRACTICE FOR TASK TWO

Skill	Grip	Cues	Ways to Practice
<b>Forehand Ground-stroke</b>	shake-hands grip eastern-forehand grip forehand grip "V" of hand lined up w/racket edge. continental western	ready position shake-hands grip side to target firm wrist watch the ball contact at front foot low to high follow-through	practice stroke w/out ball drop & hit to fence drop & hit over net hit from partner toss any activity that is listed in USTA schools program curriculum move feet to get to ball
<b>Backhand Ground-stroke</b>	eastern-backhand grip twist left knuckle on top 1/4 turn left continental shake-hands	1/4 turn inward knuckle on top side to target watch the ball low to high step across and forward a foot in front of foot point to the sky follow-through (high) contact out front	practice stroke w/out ball drop & hit to fence drop & hit over net hit from partner toss any activity that is listed in USTA schools program curriculum ball machine hit with partner
<b>Forehand Volley</b>	eastern forehand grip shake-hands grip forehand grip "V" of hand lined up with racket edge. continental	shake-hands grip firm wrist thumb to belly button side to target watch the ball contact out in front short back-swing block/punch no follow-through racket above handle	hit from partner toss hit from partner hit drop & hit over net continuous rally with partner any activity that is listed in USTA schools program curriculum ball machine pepper
<b>Backhand Volley</b>	eastern-backhand grip twist left knuckle on top 1/4 turn left continental	firm wrist side to target watch the ball step across and forward contact out in front block/punch no follow-through split-step racket above handle	hit from partner toss hit from partner hit drop & hit over net continuous rally with partner any activity that is listed in USTA schools program curriculum ball machine pepper
<b>Serve: Simplified Beginners Punch Full-swing</b>	shake-hands grip eastern forehand grip continental grip forehand grip "V" of hand lined up with racket edge.	shake-hands grip side to target palm to ear toss to the top eye on ball down together/up together swing up at the peak reach for the sky make an "X" follow-through reach for the ball	tossing straight up practice toss w/out hitting ball practice swing motion - mirror toss and hit at fence toss and hit over net actual service any activity that is listed in USTA schools program curriculum

USTA SCHOOLS PROGRAM CURRICULUM (1993)

## APPENDIX L

### RATER TRAINING: METHOD FOR REPORTING SCORES

TASK TWO:  
Sample Products from Pilot Study

<b>Task 2: Booklet #</b>	<b>Rater's Score</b>
1	4
2	1
3	4
4	4
5	1
6	2
7	3
8	2
9	3
10	2

## APPENDIX M

### RATER TRAINING: METHOD FOR REPORTING SCORES

TASK TWO:  
Rater Training - Inter-rater Reliability

<b>Task 2: Booklet</b>	<b>Rater</b>	<b>Reasearcher</b>	<b>% Reliability</b>
1	4	4	yes
2	1	1	yes
3	4	4	yes
4	4	4	yes
5	1	1	yes
6	2	2	yes
7	3	3	yes
8	2	2	yes
9	3	3	yes
10	2	2	yes
			100%

## APPENDIX N

### RATER TRAINING: METHOD FOR REPORTING SCORES

#### TASK TWO

#### Rater Training - Intra-rater Reliability

Task 2: Booklet	Rater 1st trial	Rater 2nd trial	% Reliability
1	4	4	yes
2	1	1	yes
3	4	4	yes
4	4	4	yes
5	1	1	yes
6	2	2	yes
7	3	3	yes
8	2	2	yes
9	3	3	yes
10	2	2	yes
			100%

## APPENDIX 0

### DATA COLLECTION SCHEDULE

EUVTP - Experienced University Varsity Tennis Players

EUTP - Experienced University Tennis Players

IUTP - Inexperienced University Tennis Players

SUBJECTS	TASK ONE	TRIAL ONE	SETTING	TIME ALLOTTED	TRIAL TWO
EUVTP	Creating a video	Tues., before	University courts/ gym	45 minutes	Tues., before
EUTP	Creating a video	Mon., during p.e.	University courts/gym	45 minutes	Mon., during
IUTP	Creating a video	Mon., during	University courts/gym	45 minutes	Mon., during
SUBJECTS	TASK TWO	TRIAL ONE	SETTING	TIME ALLOTTED	TRIAL TWO
EUVTP	Developing a booklet	Thurs., before	University courts/gym	45 minutes	Thurs., before
EUTP	Developing a booklet	Wed., during p.e.	University courts/gym	45 minutes	Wed., during
IUTP	Developing a booklet	Wed., during	University courts/gym	45 minutes	Wed., during

## APPENDIX P

### CONTENT VALIDITY QUESTIONS

(USTA Professional)

#### **Task One: Creating a Video of Basic Tennis Skills**

1. The content of the assessment task reflects the content of the USTA Schools Program Curriculum.

**Strongly Agree      Agree      Neither**  
**Disagree      Strongly Disagree**

**Comment:**

2. The scoring rubric is consistent with the goals and objectives of the USTA Schools Program Curriculum.

**Strongly Agree      Agree      Neither**  
**Disagree      Strongly Disagree**

**Comment:**

#### **Task Two: Developing a Booklet of Basic Tennis Skills**

1. The content of the assessment task reflects the content of the USTA Schools Program Curriculum.

**Strongly Agree      Agree      Neither**  
**Disagree      Strongly Disagree**

**Comment:**

2. The scoring rubric is consistent with the goals and objectives of the USTA Schools Program Curriculum.

**Strongly Agree      Agree      Neither**  
**Disagree      Strongly Disagree**

**Comment:**

## VITA

Karyn Mullholand Teske was born on June 20, 1969 in Hollywood, Florida. She attended private school for 12 years in Seminole, Florida, and graduated from Pinellas Park High School in 1987. She continued her education at St. Petersburg Junior college and graduated with an Associate of Arts degree. In December of 1991, she received a Bachelor of Science degree in Physical Education from Florida State University. In November of 1997, she received a Master of Science degree in Physical Education Pedagogy from Virginia Polytechnic Institute and State University. She is currently teaching Elementary Physical Education in Orlando, Florida.

---

Karyn Mullholand Teske